

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

AN

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 November 2002 (07.11.2002)

PCT

(10) International Publication Number
WO 02/088183 A2

- (51) International Patent Classification⁷: C07K 14/705,
14/72
- (21) International Application Number: PCT/GB02/01923
- (22) International Filing Date: 25 April 2002 (25.04.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
0110662.4 1 May 2001 (01.05.2001) GB
60/288,160 2 May 2001 (02.05.2001) US
- (71) Applicant (for all designated States except US): PARADIGM THERAPEUTICS LIMITED [GB/GB]; St. John's Innovation Centre, Cowley Road, Cambridge CB4 0WS (GB).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): CARLTON, Mark [GB/GB]; Avlar BioVentures, Compass House, Vision Park, Chivers Way, Histon, Cambridge CB4 9ZR (GB). DIXON, John [GB/GB]; Avlar BioVentures, Compass House, Vision Park, Chivers Way, Histon, Cambridge CB4 9ZR (GB). HENDRICK, Alan [IE/GB]; Avlar BioVentures, Compass House, Vision Park, Chivers Way, Histon, Cambridge CB4 9ZR (GB).
- (74) Agents: KHOO, Chong-Yee et al.; D Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),

[Continued on next page]

(54) Title: RECEPTOR

+ve control
-ve control
Testis
Muscle
Ovary
Prostate
Small Intestine
Lung
Kidney
Leukocytes
Liver
Brain
Heart
Spleen

(57) Abstract: We disclose Tarzan G-protein coupled receptor (GPCR) polypeptides comprising the amino acid sequence shown in SEQ ID NO: 3, SED ID NO: 5, SEQ ID NO: 6 OR SEQ ID NO: 8, and homologues, variants and derivatives thereof. Nucleic acids capable of encoding Tarzan polypeptide are also disclosed, in particular, those comprising the nucleic acid sequences shown in SEQ ID NO: 1, SEQ ID NO: 2 or SEQ ID NO: 4.

WO 02/088183 A2



Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

- *without international search report and to be republished upon receipt of that report*

RECEPTOR

FIELD

This invention relates to newly identified nucleic acids, polypeptides encoded by them and to their production and use. More particularly, the nucleic acids and polypeptides of the present invention relate to a G-protein coupled receptor (GPCR), hereinafter referred to as "Tarzan GPCR", and members of the purinoceptor family of GPCRs. The invention also relates to inhibiting or activating the action of such nucleic acids and polypeptides.

10 BACKGROUND

It is well established that many medically significant biological processes are mediated by proteins participating in signal transduction pathways that involve G-proteins and/or second messengers, for example, cAMP (Lefkowitz, *Nature*, 1991, 351: 353-354). These proteins are referred to as proteins participating in pathways with G-proteins or 15 "PPG proteins". Some examples of these proteins include the GPC receptors, such as those for adrenergic agents and dopamine (Kobilka, B. K., et al., *Proc. Natl Acad. Sci., USA*, 1987, 84: 46-50; Kobilka B. K., et al., *Science*, 1987, 238: 650-656; Bunzow, J. R., et al., *Nature*, 1988, 336: 783-787), G-proteins themselves, effector proteins, for example, phospholipase C, adenyl cyclase, and phosphodiesterase, and actuator proteins, for 20 example, protein kinase A and protein kinase C (Simon, M. I., et al., *Science*, 1991, 252: 802-8).

For example, in one form of signal transduction, the effect of hormone binding is activation of the enzyme adenylate cyclase inside the cell. Enzyme activation by hormones is dependent on the presence of the nucleotide, GTP. GTP also influences hormone 25 binding. A G-protein connects the hormone receptor to adenylate cyclase. G-protein is shown to exchange GTP for bound GDP when activated by a hormone receptor. The GTP carrying form then binds to activated adenylate cyclase. Hydrolysis of GTP to GDP, catalysed by the G-protein itself, returns the G-protein to its basal, inactive form. Thus, the

G-protein serves a dual role, as an intermediate that relays the signal from receptor to effector, and as a clock that controls the duration of the signal.

The membrane protein gene superfamily of G-protein coupled receptors (GPCRs) has been characterised as having seven putative transmembrane domains. The domains are 5 believed to represent transmembrane α -helices connected by extracellular or cytoplasmic loops. G-protein coupled receptors include a wide range of biologically active receptors, such as hormone, viral, growth factor and neuroreceptors.

G-protein coupled receptors (also known as 7TM receptors) have been characterised as including these seven conserved hydrophobic stretches of about 20 to 30 10 amino acids, connecting at least eight divergent hydrophilic loops. The G-protein family of coupled receptors includes dopamine receptors which bind to neuroleptic drugs used for treating psychotic and neurological disorders. Other examples of members of this family include, but are not limited to, calcitonin, adrenergic, endothelin, cAMP, adenosine, muscarinic, acetylcholine, serotonin, histamine, thrombin, kinin, follicle stimulating 15 hormone, opsins, endothelial differentiation gene-1, rhodopsins, odorant, and cytomegalovirus receptors.

Most G-protein coupled receptors have single conserved cysteine residues in each of the first two extracellular loops which form disulphide bonds that are believed to stabilise functional protein structure. The 7 transmembrane regions are designated as TM1, 20 TM2, TM3, TM4, TM5, TM6, and TM7. TM3 has been implicated in signal transduction.

Phosphorylation and lipidation (pamitylation or farnesylation) of cysteine residues can influence signal transduction of some G-protein coupled receptors. Most G-protein coupled receptors contain potential phosphorylation sites within the third cytoplasmic loop and/or the carboxy terminus. For several G-protein coupled receptors, such as the β - 25 adrenoreceptor, phosphorylation by protein kinase A and/or specific receptor kinases mediates receptor desensitization. For some receptors, the ligand binding sites of G-protein coupled receptors are believed to comprise hydrophilic sockets formed by several G-protein coupled receptor transmembrane domains, the sockets being surrounded by

hydrophobic residues of the G-protein coupled receptors. The hydrophilic side of each G-protein coupled receptor transmembrane helix is thought to face inward and form a polar ligand binding site. TM3 has been implicated in several G-protein coupled receptors as having a ligand binding site, such as the TM3 aspartate residue. TM5 serines, a TM6 asparagine and TM6 or TM7 phenylalanines or tyrosines are also implicated in ligand binding.

G-protein coupled receptors can be intracellularly coupled by heterotrimeric G-proteins to various intracellular enzymes, ion channels and transporters (see, Johnson et al., *Endoc. Rev.*, 1989, 10: 317-331). Different G-protein α -subunits preferentially stimulate particular effectors to modulate various biological functions in a cell. Phosphorylation of cytoplasmic residues of G-protein coupled receptors has been identified as an important mechanism for the regulation of G-protein coupling of some G-protein coupled receptors. G-protein coupled receptors are found in numerous sites within a mammalian host. Over the past 15 years, nearly 350 therapeutic agents targeting 7 transmembrane (7 TM) receptors have been successfully introduced onto the market.

Thus, G-protein coupled receptors have an established, proven history as therapeutic targets. Clearly there is a need for identification and characterization of further receptors which can play a role in preventing, ameliorating or correcting dysfunctions or diseases, including, but not limited to, infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; diabetes, obesity; anorexia; bulimia; asthma; parkinson's disease; thrombosis; acute heart failure; hypotension; hypertension; erectile dysfunction; urinary retention; metabolic bone diseases such as osteoporosis and osteo petrosis; angina pectoris; myocardial infarction; ulcers; asthma; allergies; rheumatoid arthritis; inflammatory bowel disease; irritable bowel syndrome benign prostatic hypertrophy; and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles dela Tourett's syndrome.

SUMMARY OF THE INVENTION

According to a first aspect of the present invention, we provide a Tarzan GPCR polypeptide comprising the amino acid sequence shown in SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 6 or SEQ ID NO: 8, or a homologue, variant or derivative thereof.

5 There is provided, according to a second aspect of the present invention, a nucleic acid capable of encoding a polypeptide according to the first aspect of the invention. Preferably, the nucleic acid comprises the nucleic acid sequence shown in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 7 or SEQ ID NO: 10, or a homologue, variant or derivative thereof.

10 We provide, according to a third aspect of the present invention, a polypeptide comprising a fragment of a polypeptide according to the first aspect of the invention.

Preferably, such a fragment containing polypeptide comprises one or more regions which are homologous between any two or more of SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 6 and SEQ ID NO: 8, or which comprises one or more regions which are
15 heterologous between any two or more of SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 6 and SEQ ID NO: 8.

As a fourth aspect of the present invention, there is provided a nucleic acid capable of encoding a polypeptide according to the third aspect of the invention.

20 We provide, according to a fifth aspect of the present invention, a vector comprising a nucleic acid according to the second or fourth aspect of the invention.

The present invention, in a sixth aspect, provides a host cell comprising a nucleic acid according to the second or fourth aspect of the invention, or vector according to the fifth aspect of the invention.

In a seventh aspect of the present invention, there is provided a transgenic non-human animal comprising a nucleic acid according to the second or fourth aspect of the invention or a vector according to the fifth aspect of the invention. Preferably, the transgenic non-human animal is a mouse.

5 According to an eighth aspect of the present invention, we provide use of a polypeptide according to the first or third aspect of the invention in a method of identifying compound which is capable of interacting specifically with a G protein coupled receptor.

We provide, according to a ninth aspect of the invention, use of a transgenic non-
10 human animal according to the seventh aspect of the invention in a method of identifying a compound which is capable of interacting specifically with a G protein coupled receptor.

There is provided, in accordance with a tenth aspect of the present invention, a method for identifying an antagonist of a Tarzan GPCR, the method comprising contacting
a cell which expresses Tarzan receptor with a candidate compound and determining
15 whether the level of cyclic AMP (cAMP) in said cell is lowered as a result of said contacting.

As an eleventh aspect of the invention, we provide a method for identifying a compound capable of lowering the endogenous level of cyclic AMP in a cell which method comprises contacting a cell which expresses a Tarzan GPCR with a candidate
20 compound and determining whether the level of cyclic AMP (cAMP) in said cell is lowered as a result of said contacting.

According to a twelfth aspect of the invention, we provide a method for identifying a compound capable of binding to a Tarzan GPCR polypeptide, the method comprising contacting a Tarzan GPCR polypeptide with a candidate compound and determining
25 whether the candidate compound binds to the Tarzan GPCR polypeptide.

We provide, according to a thirteenth aspect of the invention, there is provided a compound identified by a method according to any of the eighth to twelfth aspects of the invention.

According to a fourteenth aspect of the present invention, we provide a compound
5 capable of binding specifically to a polypeptide according to the first or third aspect of the invention.

There is provided, according to a fifteenth aspect of the present invention, use of a polypeptide according to the first or third aspect of the invention, or part thereof; or a nucleic acid according to the second or fourth aspect of the invention, or part thereof, in a
10 method for producing antibodies.

We provide, according to a sixteenth aspect of the present invention, an antibody capable of binding specifically to a polypeptide according to the first or third aspect of the invention, or part thereof; or a polypeptide encoded by a nucleic acid according to the second or fourth aspect of the invention, or part thereof.,

15 As a seventeenth aspect of the present invention, there is provided a pharmaceutical composition comprising any one or more of the following: a polypeptide according to the first or third aspect of the invention, or part thereof; a polypeptide encoded by a nucleic acid according to the second or fourth aspect of the invention, or part thereof; a vector according to the fifth aspect of the invention; a cell according to the sixth aspect of the invention; a compound according to the thirteenth or fourteenth aspect of the invention; and an antibody according to the sixteenth aspect of the invention, together with
20 a pharmaceutically acceptable carrier or diluent.

We provide, according to a eighteenth aspect of the present invention, a vaccine composition comprising any one or more of the following: a polypeptide according to the first or third aspect of the invention, or part thereof; a polypeptide encoded by a nucleic acid according to the second or fourth aspect of the invention, or part thereof; a vector according to the fifth aspect of the invention; a cell according to the sixth aspect of the

invention; a compound according to the thirteenth or fourteenth aspect of the invention; and an antibody according to the sixteenth aspect of the invention.

- According to an nineteenth aspect of the present invention, we provide a diagnostic kit for a disease or susceptibility to a disease comprising any one or more of the following:
- 5 a polypeptide according to the first or third aspect of the invention, or part thereof; a polypeptide encoded by a nucleic acid according to the second or fourth aspect of the invention, or part thereof; a vector according to the fifth aspect of the invention; a cell according to the sixth aspect of the invention; a compound according to the thirteenth or fourteenth aspect of the invention; and an antibody according to the sixteenth aspect of the
- 10 invention.

We provide, according to a twentieth aspect of the invention, a method of treating a patient suffering from a disease associated with enhanced activity of a Tarzan GPCR, which method comprises administering to the patient an antagonist of Tarzan GPCR.

- There is provided, in accordance with a twenty-first aspect of the present
- 15 invention, a method of treating a patient suffering from a disease associated with reduced activity of a Tarzan GPCR, which method comprises administering to the patient an agonist of Tarzan GPCR.

Preferably, the Tarzan GPCR comprises a polypeptide having the sequence shown in SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 6 or SEQ ID NO: 8.

- 20 According to a twenty-second aspect of the present invention, we provide a method for treating and/or preventing a disease in a patient, which comprises the step of administering any one or more of the following to the patient: a polypeptide according to the first or third aspect of the invention, or part thereof; a polypeptide encoded by a nucleic acid according to the second or fourth aspect of the invention, or part thereof; a vector according to the fifth aspect of the invention; a cell according to the sixth aspect of the invention; a compound according to the thirteenth or fourteenth aspect of the invention; and an antibody according to the sixteenth aspect of the invention; a
- 25

pharmaceutical composition according to the seventeenth aspect of the invention; and a vaccine according to the eighteenth aspect of the invention, to the subject.

There is provided, according to a twenty-third aspect of the present invention, an agent comprising a polypeptide according to the first or third aspect of the invention, or part thereof; a polypeptide encoded by a nucleic acid according to the second or fourth aspect of the invention, or part thereof; a vector according to the fifth aspect of the invention; a cell according to the sixth aspect of the invention; a compound according to the thirteenth or fourteenth aspect of the invention; and an antibody according to the sixteenth aspect of the invention, said agent for use in a method of treatment or prophylaxis of disease.

We provide, according to a twenty-fourth aspect of the present invention, use of a polypeptide according to the first or third aspect of the invention, or part thereof; a polypeptide encoded by a nucleic acid according to the second or fourth aspect of the invention, or part thereof; a vector according to the fifth aspect of the invention; a cell according to the sixth aspect of the invention; a compound according to the thirteenth or fourteenth aspect of the invention; and an antibody according to the sixteenth aspect of the invention, for the preparation of a pharmaceutical composition for the treatment or prophylaxis of a disease.

As a twenty-fifth aspect of the present invention, there is provided non-human transgenic animal, characterized in that the transgenic animal comprises an altered Tarzan gene. Preferably, the alteration is selected from the group consisting of: a deletion of Tarzan, a mutation in Tarzan resulting in loss of function, introduction of an exogenous gene having a nucleotide sequence with targeted or random mutations into Tarzan, introduction of an exogenous gene from another species into Tarzan, and a combination of any of these.

We provide, according to a twenty-sixth aspect of the present invention, a non-human transgenic animal having a functionally disrupted endogenous Tarzan gene, in

which the transgenic animal comprises in its genome and expresses a transgene encoding a heterologous Tarzan protein.

The present invention, in a twenty-seventh aspect, provides a nucleic acid construct for functionally disrupting a Tarzan gene in a host cell, the nucleic acid construct comprising: (a) a non-homologous replacement portion; (b) a first homology region located upstream of the non-homologous replacement portion, the first homology region having a nucleotide sequence with substantial identity to a first Tarzan gene sequence; and (c) a second homology region located downstream of the non-homologous replacement portion, the second homology region having a nucleotide sequence with substantial identity to a second Tarzan gene sequence, the second Tarzan gene sequence having a location downstream of the first Tarzan gene sequence in a naturally occurring endogenous Tarzan gene.

According to a twenty-eighth aspect of the present invention, we provide a process for producing a Tarzan GPCR polypeptide, the method comprising culturing a host cell according to the sixth aspect of the invention under conditions in which a nucleic acid encoding a Tarzan GPCR polypeptide is expressed.

There is provided, according to a twenty-ninth aspect of the present invention, a method of detecting the presence of a nucleic acid according to the second or fourth aspect of the invention in a sample, the method comprising contacting the sample with at least one nucleic acid probe which is specific for said nucleic acid and monitoring said sample for the presence of the nucleic acid.

We provide, according to a thirtieth aspect of the present invention, a method of detecting the presence of a polypeptide according to the first or third aspect of the invention in a sample, the method comprising contacting the sample with an antibody according to the sixteenth aspect of the invention and monitoring said sample for the presence of the polypeptide.

As a thirty-first aspect of the present invention, there is provided a method of diagnosis of a disease or syndrome caused by or associated with increased, decreased or otherwise abnormal expression of Tarzan GPCR, the method comprising the steps of: (a) detecting the level or pattern of expression of Tarzan GPCR in an animal suffering or suspected to be suffering from such a disease; and (b) comparing the level or pattern of expression with that of a normal animal.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a diagram showing the results of analysis of the human Tarzan polypeptide (SEQ ID NO: 3) using the HMM structural prediction software of pfam (http://www.sanger.ac.uk/Software/Pfam/search.shtml).

Figure 2 is a diagram showing an expression profile for human Tarzan GPCR generated by reverse transcription-polymerase chain reaction (RT-PCR).

Figure 3 is a diagram showing the structure of the mouse Tarzan locus before knock-out.

Figure 4 is a diagram showing the structure of the mouse Tarzan locus after knock-out.

Figure 5 is a diagram showing the structure of the vector used for mouse Tarzan knock-out, including cloning sites.

SEQUENCE LISTINGS

SEQ ID NO: 1 shows the cDNA sequence of human Tarzan. SEQ ID NO: 2 shows an open reading frame derived from SEQ ID NO: 1. SEQ ID NO: 3 shows the amino acid sequence of human Tarzan. SEQ ID NO: 4 shows the open reading frame of a cDNA for mouse Tarzan. SEQ ID NO: 5 shows the amino acid sequence of mouse Tarzan.

SEQ ID NO: 6 shows the amino acid sequence of a fusion protein of TARZAN with V5 and His tags at c-terminal end. **SEQ ID NO: 7** shows a nucleotide sequence comprising Tarzan with NheI and XhoI at the 5-prime and 3-prime ends, generated by PCR. **SEQ ID NO: 8** shows the amino acid sequence of Tarzan fusion protein comprising 5 a C terminal FLAG tag. **SEQ ID NO: 10** shows the nucleotide sequence of a Tarzan genomic locus from 5'prF to 3'prR, including translated coding amino acid sequence.

Where two sequences are set out under a particular sequence identifier (i.e., SEQ ID NO: NN), both sequences should understood to be encompassed by reference to that particular SEQ ID NO:.

10 **DETAILED DESCRIPTION**

TARZAN GPCR

Our invention relates in general to a novel G-Protein Coupled Receptor (GPCR), in particular, an orphan purinoceptor type G-protein coupled receptor, which we refer to as Tarzan GPCR, as well as homologues, variants or derivatives thereof.

15 Tarzan is structurally related to other proteins of the G-protein coupled receptor family, as shown by the results of sequencing the amplified cDNA products encoding human Tarzan. The cDNA sequence of SEQ ID NO: 1 contains an open reading flame (SEQ ID NO: 2, nucleotide numbers 11 to 1022) encoding a polypeptide of 337 amino acids shown in SEQ ID NO: 3. Human Tarzan is found to map to homo sapiens
20 chromosome 13q32.2.

IDENTITIES AND SIMILARITIES TO TARZAN

The amino acid sequence of Tarzan has about 36 % identity and 58 % similarity (using BLAST) in 299 amino acid residues with human p2y purinoceptor 1 (p2y1) (accession # CAA07339.1, Leon, C., Vial, C., Weber, J., Cazenave, J.P. and Gacher, C.
25 direct Submission (15-JUN-1998) related accession numbers: NP_002554.1, P47900,

JC4737, AAA97872.1, AAA97873.1, AAB47091.1). The nucleotide sequence of Tarzan (SEQ ID NO: 1) has about 100% identity (using BLAST) in 178 nucleotide residues with the anonymous homo sapiens EST normal thyroid epithelium cdna from accession # AW827323 Robert Strausberg, NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>, May 18
5 2000.

Analysis of the Tarzan polypeptide (SEQ ID NO: 3) using the HMM structural prediction software of pfam (<http://www.sanger.ac.uk/Software/Pfam/search.shtml>) confirms that Tarzan peptide is a GPCR of the 7TM-1 structural class (see Figure 1).

Human and mouse Tarzan GPCR are therefore members of a large family of G
10 Protein Coupled Receptors (GPCRs).

EXPRESSION PROFILE OF TARZAN

Polymerase chain reaction (PCR) amplification of Tarzan cDNA detects expression of Tarzan to varying abundance in human spleen, heart, brain, kidney, lung, small intestine, prostate, ovary, muscle, and testis. An expression profile of Tarzan GPCR
15 is shown in Figure 2. Using Tarzan cDNA of SEQ ID NO: 1 to search the human EST data sources by BLASTN, identities are found in cDNA derived from libraries originating from normal thyroid epithelium (accession # AW827323). This indicates that Tarzan is expressed in these normal or abnormal tissues. Accordingly, the Tarzan polypeptides, nucleic acids, probes, antibodies, expression vectors and ligands are useful for detection,
20 diagnosis, treatment and other assays for diseases associated with over-, under- and abnormal expression of Tarzan GPCR in these and other tissues.

Furthermore, the Tarzan polypeptides, nucleic acids, probes, antibodies, expression vectors and ligands are useful as (and for the production of) biosensors. According to Aizawa (1988), *Anal. Chem. Symp.* 17: 683, a biosensor is defined as being a unique
25 combination of a receptor for molecular recognition, for example a selective layer with immobilized antibodies or receptors such as a Tarzan G-protein coupled receptor, and a transducer for transmitting the values measured. One group of such biosensors will detect

the change which is caused in the optical properties of a surface layer due to the interaction of the receptor with the surrounding medium. Among such techniques may be mentioned especially ellipso-metry and surface plasmon resonance. Biosensors incorporating Tarzan may be used to detect the presence or level of Tarzan ligands, for example, nucleotides such as purines or purine analogues, or analogues of these ligands. The construction of such biosensors is well known in the art. Thus, cell lines expressing Tarzan receptor may be used as reporter systems for detection of ligands such as ATP via receptor-promoted formation of [3H]inositol phosphates or other second messengers (Watt et al., 1998, *J Biol Chem* May 29;273(22):14053-8). Receptor-ligand biosensors are also described in Hoffman et al., 2000, *Proc Natl Acad Sci U S A* Oct 10;97(21):11215-20. Optical and other biosensors comprising Tarzan may also be used to detect the level or presence of interaction with G-proteins and other proteins, as described by, for example, Figler et al, 1997, *Biochemistry* Dec 23;36(51):16288-99 and Sarrio et al., 2000, *Mol Cell Biol* 2000 Jul;20(14):5164-74). Sensor units for biosensors are described in, for example, US 5,492,840.

This and other embodiments of the invention will be described in further detail below.

METHODS EMPLOYED

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA and immunology, which are within the capabilities of a person of ordinary skill in the art. Such techniques are explained in the literature. See, for example, J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Books 1-3, Cold Spring Harbor Laboratory Press; Ausubel, F. M. et al. (1995 and periodic supplements; *Current Protocols in Molecular Biology*, ch. 9, 13, and 16, John Wiley & Sons, New York, N.Y.); B. Roe, J. Crabtree, and A. Kahn, 1996, *DNA Isolation and Sequencing: Essential Techniques*, John Wiley & Sons; J. M. Polak and James O'D. McGee, 1990, *In Situ Hybridization: Principles and Practice*; Oxford University Press; M. J. Gait (Editor), 1984, *Oligonucleotide Synthesis: A Practical Approach*, Irl Press; and, D.

M. J. Lilley and J. E. Dahlberg, 1992, *Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA Methods in Enzymology*, Academic Press. Each of these general texts is herein incorporated by reference.

TARZAN GPCR POLYPEPTIDES

5 As used here, the term "Tarzan GPCR polypeptide" is intended to refer to a polypeptide comprising the amino acid sequence shown in SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 6 or SEQ ID NO: 8, or a homologue, variant or derivative thereof. Preferably, the polypeptide comprises or is a homologue, variant or derivative of the sequence shown in SEQ ID NO: 3.

10 "Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids.

15 "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications.

20 Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a

- nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. See, for instance, *Proteins - Structure and Molecular Properties*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993 and Wold, F., 5 *Posttranslational Protein Modifications: Perspectives and Prospects*, pgs. 1-12 in *Posttranslational Covalent Modification of Proteins*, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter et al., "Analysis for protein modifications and nonprotein cofactors", *Meth Enzymol* (1990) 182:626-646 and Rattan et al, "Protein Synthesis: Posttranslational Modifications and Aging", *Ann NY Acad Sci* (1992) 663:48-62.
- 10 15 The terms "variant", "homologue", "derivative" or "fragment" in relation to the present invention include any substitution of, variation of, modification of, replacement of, deletion of or addition of one (or more) amino acid from or to a sequence. Unless the context admits otherwise, references to "Tarzan" and "Tarzan GPCR" include references to such variants, homologues, derivatives and fragments of Tarzan.
- 20 25 Preferably, as applied to Tarzan, the resultant amino acid sequence has GPCR activity, more preferably having at least the same activity of the Tarzan GPCR shown as SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 6 or SEQ ID NO: 8. In particular, the term "homologue" covers identity with respect to structure and/or function providing the resultant amino acid sequence has GPCR activity. With respect to sequence identity (i.e. similarity), preferably there is at least 70%, more preferably at least 75%, more preferably at least 85%, even more preferably at least 90% sequence identity. More preferably there is at least 95%, more preferably at least 98%, sequence identity. These terms also encompass polypeptides derived from amino acids which are allelic variations of the Tarzan GPCR nucleic acid sequence.

- Where reference is made to the "receptor activity" or "biological activity" of a receptor such as Tarzan GPCR, these terms are intended to refer to the metabolic or physiological function of the Tarzan receptor, including similar activities or improved activities or these activities with decreased undesirable side effects. Also included are
- 5 antigenic and immunogenic activities of the Tarzan receptor. Examples of GPCR activity, and methods of assaying and quantifying these activities, are known in the art, and are described in detail elsewhere in this document.

As used herein a "deletion" is defined as a change in either nucleotide or amino acid sequence in which one or more nucleotides or amino acid residues, respectively, are

10 absent. As used herein an "insertion" or "addition" is that change in a nucleotide or amino acid sequence which has resulted in the addition of one or more nucleotides or amino acid residues, respectively, as compared to the naturally occurring substance. As used herein "substitution" results from the replacement of one or more nucleotides or amino acids by different nucleotides or amino acids, respectively.

15 Tarzan polypeptides according to the present invention may also have deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent amino acid sequence. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine, glycine, alanine, asparagine, glutamine, serine, threonine, phenylalanine, and tyrosine.

20

Conservative substitutions may be made, for example according to the table below.

25 Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q

	Polar - charged	D E
		K R
AROMATIC		H F W Y

- Tarzan polypeptides of the invention may further comprise heterologous amino acid sequences, typically at the N-terminus or C-terminus, preferably the N-terminus. Heterologous sequences may include sequences that affect intra or extracellular protein targeting (such as leader sequences). Heterologous sequences may also include sequences
- 5 that increase the immunogenicity of the polypeptide of the invention and/or which facilitate identification, extraction and/or purification of the polypeptides. Another heterologous sequence that is particularly preferred is a polyamino acid sequence such as polyhistidine which is preferably N-terminal. A polyhistidine sequence of at least 10 amino acids, preferably at least 17 amino acids but fewer than 50 amino acids is especially
- 10 preferred.

- The Tarzan GPCR polypeptides may be in the form of the "mature" protein or may be a part of a larger protein such as a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an
- 15 additional sequence for stability during recombinant production.

- Tarzan polypeptides of the invention are advantageously made by recombinant means, using known techniques. However they may also be made by synthetic means using techniques well known to skilled persons such as solid phase synthesis. Polypeptides of the invention may also be produced as fusion proteins, for example to aid in extraction and purification. Examples of fusion protein partners include glutathione-S-transferase (GST), 6xHis, GAL4 (DNA binding and/or transcriptional activation domains) and β -galactosidase. It may also be convenient to include a proteolytic cleavage site between the fusion protein partner and the protein sequence of interest to allow removal of fusion protein sequences, such as a thrombin cleavage site. Preferably the fusion protein will not hinder the function of the protein of interest sequence.

Tarzan polypeptides of the invention may be in a substantially isolated form. This term is intended to refer to alteration by the hand of man from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide, nucleic acid or a 5 polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide, nucleic acid or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

It will however be understood that the Tarzan GPCR protein may be mixed with carriers or diluents which will not interfere with the intended purpose of the protein and 10 still be regarded as substantially isolated. A polypeptide of the invention may also be in a substantially purified form, in which case it will generally comprise the protein in a preparation in which more than 90%, for example, 95%, 98% or 99% of the protein in the preparation is a Tarzan GPCR polypeptide of the invention.

The present invention also relates to peptides comprising a portion of a Tarzan 15 polypeptide according to the invention. Thus, fragments of Tarzan GPCR and its homologues, variants or derivatives are included. The peptides of the present invention may be between 2 and 200 amino acids, preferably between 4 and 40 amino acids in length. The peptide may be derived from a Tarzan GPCR polypeptide as disclosed here, for example by digestion with a suitable enzyme, such as trypsin. Alternatively the 20 peptide, fragment, etc may be made by recombinant means, or synthesised synthetically,

The term "peptide" includes the various synthetic peptide variations known in the art, such as a retroinverso D peptides. The peptide may be an antigenic determinant and/or a T-cell epitope. The peptide may be immunogenic *in vivo*. Preferably the peptide is capable of inducing neutralising antibodies *in vivo*.

25 By aligning Tarzan GPCR sequences from different species, it is possible to determine which regions of the amino acid sequence are conserved between different species ("homologous regions"), and which regions vary between the different species ("heterologous regions").

The Tarzan polypeptides according to the invention may therefore comprise a sequence which corresponds to at least part of a homologous region. A homologous region shows a high degree of homology between at least two species. For example, the homologous region may show at least 70%, preferably at least 80%, more preferably at 5 least 90%, even more preferably at least 95% identity at the amino acid level using the tests described above. Peptides which comprise a sequence which corresponds to a homologous region may be used in therapeutic strategies as explained in further detail below. Alternatively, the Tarzan GPCR peptide may comprise a sequence which corresponds to at least part of a heterologous region. A heterologous region shows a low 10 degree of homology between at least two species.

TARZAN GPCR POLYNUCLEOTIDES AND NUCLEIC ACIDS

This invention encompasses Tarzan polynucleotides, Tarzan nucleotides and Tarzan nucleic acids, methods of production, uses of these, etc, as described in further detail elsewhere in this document.

15 The terms "Tarzan polynucleotide", "Tarzan nucleotide" and "Tarzan nucleic acid" may be used interchangeably, and are intended to refer to a polynucleotide/nucleic acid comprising a nucleic acid sequence as shown in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 7 or SEQ ID NO: 10, or a homologue, variant or derivative thereof. Preferably, the polynucleotide/nucleic acid comprises or is a homologue, variant or 20 derivative of the nucleic acid sequence SEQ ID NO: 1 or SEQ ID NO: 2, most preferably, SEQ ID NO: 2.

These terms are also intended to include a nucleic acid sequence capable of encoding a polypeptides and/or a peptide of the present invention, i.e., a Tarzan polypeptide. Thus, Tarzan GPCR polynucleotides and nucleic acids comprise a nucleotide 25 sequence capable of encoding a polypeptide comprising the amino acid sequence shown in SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 6 or SEQ ID NO: 8, or a homologue, variant or derivative thereof. Preferably, the Tarzan GPCR polynucleotides and nucleic acids

comprise a nucleotide sequence capable of encoding a polypeptide comprising the amino acid sequence shown in SEQ ID NO: 3, or a homologue, variant or derivative thereof.

- “Polynucleotide” generally refers to any polyribonucleotide or polydeoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. “Polynucleotides” include, without limitation single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, “polynucleotide” refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. “Modified” bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications has been made to DNA and RNA; thus, “polynucleotide” embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. “Polynucleotide” also embraces relatively short polynucleotides, often referred to as oligonucleotides.

- It will be understood by the skilled person that numerous nucleotide sequences can encode the same polypeptide as a result of the degeneracy of the genetic code.

- As used herein, the term “nucleotide sequence” refers to nucleotide sequences, oligonucleotide sequences, polynucleotide sequences and variants, homologues, fragments and derivatives thereof (such as portions thereof). The nucleotide sequence may be DNA or RNA of genomic or synthetic or recombinant origin which may be double-stranded or single-stranded whether representing the sense or antisense strand or combinations thereof. The term nucleotide sequence may be prepared by use of recombinant DNA techniques (for example, recombinant DNA).

Preferably, the term “nucleotide sequence” means DNA.

- The terms "variant", "homologue", "derivative" or "fragment" in relation to the present invention include any substitution of, variation of, modification of, replacement of, deletion of or addition of one (or more) nucleic acids from or to the sequence of a Tarzan nucleotide sequence. Unless the context admits otherwise, references to "Tarzan" and
- 5 "Tarzan GPCR" include references to such variants, homologues, derivatives and fragments of Tarzan.

- Preferably, the resultant nucleotide sequence encodes a polypeptide having GPCR activity, preferably having at least the same activity of the GPCR shown as SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 6 or SEQ ID NO: 8. Preferably, the term "homologue" is
- 10 intended to cover identity with respect to structure and/or function such that the resultant nucleotide sequence encodes a polypeptide which has GPCR activity. With respect to sequence identity (i.e. similarity), preferably there is at least 70%, more preferably at least 75%, more preferably at least 85%, more preferably at least 90% sequence identity. More preferably there is at least 95%, more preferably at least 98%, sequence identity. These
- 15 terms also encompass allelic variations of the sequences.

TARZAN CONTROL REGIONS

- For some purposes, it may be necessary to utilise or investigate control regions of Tarzan. Such control regions include promoters, enhancers and locus control regions. By a control region we mean a nucleic acid sequence or structure which is capable of
- 20 modulating the expression of a coding sequence which is operatively linked to it.

- For example, control regions are useful in generating transgenic animals expressing Tarzan. Furthermore, control regions may be used to generate expression constructs for Tarzan. Control regions from different individuals in a population may be sequenced to identify non-coding polymorphisms, which may affect the expression level
- 25 of Tarzan. This is described in further detail below.

Identification of control regions of Tarzan is straightforward, and may be carried out in a number of ways. For example, the coding sequence of Tarzan may be obtained

from an organism, by screening a cDNA library using a human or mouse Tarzan cDNA sequence as a probe. 5' sequences may be obtained by screening an appropriate genomic library, or by primer extension as known in the art. Database searching of genome databases may also be employed. Such 5' sequences which are particularly of interest 5 include non-coding regions. The 5' regions may be examined by eye, or with the aid of computer programs, to identify sequence motifs which indicate the presence of promoter and/or enhancer regions.

Furthermore, sequence alignments may be conducted of Tarzan nucleic acid sequences from two or more organisms. By aligning Tarzan GPCR sequences from 10 different species, it is possible to determine which regions of the nucleotide sequence are conserved between different species. Such conserved regions are likely to contain control regions for the gene in question (i.e., Tarzan). The mouse and human genomic sequences as disclosed here, for example, a mouse Tarzan genomic sequence (SEQ ID NO: 10), may be employed for this purpose. Furthermore, Tarzan homologues from other organisms may 15 be obtained using standard methods of screening using appropriate probes generated from the mouse and human Tarzan sequences. The genome of the pufferfish (*Takifugu rubripes*) may also be screened to identify a Tarzan homologue; comparison of the 5' non-coding region of the Fugu Tarzan gene with a mouse or human genomic Tarzan sequence (e.g., SEQ ID NO: 10, mouse Tarzan genomic sequence) may be used to identify conserved 20 regions containing control regions.

Deletion studies may also be conducted to identify promoter and/or enhancer regions for Tarzan.

The identity of putative control regions may be confirmed by molecular biology experiments, in which the candidate sequences are linked to a reporter gene and the 25 expression of the reporter detected.

CALCULATION OF SEQUENCE HOMOLOGY

Sequence identity with respect to any of the sequences presented here can be determined by a simple "eyeball" comparison (i.e. a strict comparison) of any one or more of the sequences with another sequence to see if that other sequence has, for example, at least 70% sequence identity to the sequence(s).

Relative sequence identity can also be determined by commercially available computer programs that can calculate % identity between two or more sequences using any suitable algorithm for determining identity, using for example default parameters. A typical example of such a computer program is CLUSTAL. Other computer program methods to determine identify and similarity between the two sequences include but are not limited to the GCG program package (Devereux *et al* 1984 Nucleic Acids Research 12: 387) and FASTA (Atschul *et al* 1990 J Molec Biol 403-410).

% homology may be calculated over contiguous sequences, i.e. one sequence is aligned with the other sequence and each amino acid in one sequence is directly compared with the corresponding amino acid in the other sequence, one residue at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues.

Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion will cause the following amino acid residues to be put out of alignment, thus potentially resulting in a large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without penalising unduly the overall homology score. This is achieved by inserting "gaps" in the sequence alignment to try to maximise local homology.

However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment

- with as few gaps as possible - reflecting higher relatedness between the two compared sequences - will achieve a higher score than one with many gaps. "Affine gap costs" are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties will of course produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example, when using the GCG Wisconsin Bestfit package the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension.
- 10 Calculation of maximum % homology therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (University of Wisconsin, U.S.A.; Devereux *et al.*, 1984, Nucleic Acids Research 12:387). Examples of other software than can perform sequence comparisons include, but are not limited to, the
- 15 BLAST package (Ausubel *et al.*, 1999 *ibid* – Chapter 18), FASTA (Atschul *et al.*, 1990, J. Mol. Biol., 403-410) and the GENWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (Ausubel *et al.*, 1999 *ibid*, pages 7-58 to 7-60).

Although the final % homology can be measured in terms of identity, the
20 alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix - the default matrix for the BLAST suite of programs. GCG Wisconsin programs generally use either the public
25 default values or a custom symbol comparison table if supplied. It is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix, such as BLOSUM62.

Advantageously, the BLAST algorithm is employed, with parameters set to default values. The BLAST algorithm is described in detail at

http://www.ncbi.nih.gov/BLAST/blast_help.html, which is incorporated herein by reference. The search parameters are defined as follows, can be advantageously set to the defined default parameters.

Advantageously, "substantial identity" when assessed by BLAST equates to
5 sequences which match with an EXPECT value of at least about 7, preferably at least about 9 and most preferably 10 or more. The default threshold for EXPECT in BLAST searching is usually 10.

BLAST (Basic Local Alignment Search Tool) is the heuristic search algorithm employed by the programs blastp, blastn, blastx, tblastn, and tblastx; these programs
10 ascribe significance to their findings using the statistical methods of Karlin and Altschul (Karlin and Altschul 1990, *Proc. Natl. Acad. Sci. USA* 87:2264-68; Karlin and Altschul, 1993, *Proc. Natl. Acad. Sci. USA* 90:5873-7; see
15 http://www.ncbi.nih.gov/BLAST/blast_help.html) with a few enhancements. The BLAST programs are tailored for sequence similarity searching, for example to identify homologues to a query sequence. For a discussion of basic issues in similarity searching of sequence databases, see Altschul *et al* (1994) *Nature Genetics* 6:119-129.

The five BLAST programs available at <http://www.ncbi.nlm.nih.gov> perform the following tasks: **blastp** - compares an amino acid query sequence against a protein sequence database; **blastn** - compares a nucleotide query sequence against a nucleotide sequence database; **blastx** - compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database; **tblastn** - compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands); **tblastx** - compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.
25

BLAST uses the following search parameters:

HISTOGRAM - Display a histogram of scores for each search; default is yes. (See parameter H in the BLAST Manual).

DESCRIPTIONS - Restricts the number of short descriptions of matching sequences reported to the number specified; default limit is 100 descriptions. (See parameter V in the manual page).

EXPECT - The statistical significance threshold for reporting matches against database sequences; the default value is 10, such that 10 matches are expected to be found merely by chance, according to the stochastic model of Karlin and Altschul (1990). If the statistical significance ascribed to a match is greater than the EXPECT threshold, the match will not be reported. Lower EXPECT thresholds are more stringent, leading to fewer chance matches being reported. Fractional values are acceptable. (See parameter E in the BLAST Manual).

CUTOFF - Cutoff score for reporting high-scoring segment pairs. The default value is calculated from the EXPECT value (see above). HSPs are reported for a database sequence only if the statistical significance ascribed to them is at least as high as would be ascribed to a lone HSP having a score equal to the CUTOFF value. Higher CUTOFF values are more stringent, leading to fewer chance matches being reported. (See parameter S in the BLAST Manual). Typically, significance thresholds can be more intuitively managed using EXPECT.

ALIGNMENTS - Restricts database sequences to the number specified for which high-scoring segment pairs (HSPs) are reported; the default limit is 50. If more database sequences than this happen to satisfy the statistical significance threshold for reporting (see EXPECT and CUTOFF below), only the matches ascribed the greatest statistical significance are reported. (See parameter B in the BLAST Manual).

MATRIX - Specify an alternate scoring matrix for BLASTP, BLASTX, TBLASTN and TBLASTX. The default matrix is BLOSUM62 (Henikoff & Henikoff, 1992). The valid alternative choices include: PAM40, PAM120, PAM250 and

IDENTITY. No alternate scoring matrices are available for BLASTN; specifying the MATRIX directive in BLASTN requests returns an error response.

- STRAND - Restrict a TBLASTN search to just the top or bottom strand of the database sequences; or restrict a BLASTN, BLASTX or TBLASTX search to just reading 5 frames on the top or bottom strand of the query sequence.

- FILTER - Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (1993) Computers and Chemistry 17:149-163, or segments consisting of short-periodicity internal repeats, as determined by the XNU program of Claverie & States (1993) Computers and 10 Chemistry 17:191-201, or, for BLASTN, by the DUST program of Tatusov and Lipman (see <http://www.ncbi.nlm.nih.gov>). Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (e.g., hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

- 15 Low complexity sequence found by a filter program is substituted using the letter "N" in nucleotide sequence (e.g., "NNNNNNNNNNNNNN") and the letter "X" in protein sequences (e.g., "XXXXXXXXXX").

Filtering is only applied to the query sequence (or its translation products), not to database sequences. Default filtering is DUST for BLASTN, SEG for other programs.

- 20 It is not unusual for nothing at all to be masked by SEG, XNU, or both, when applied to sequences in SWISS-PROT, so filtering should not be expected to always yield an effect. Furthermore, in some cases, sequences are masked in their entirety, indicating that the statistical significance of any matches reported against the unfiltered query sequence should be suspect.

- 25 NCBI-gi - Causes NCBI gi identifiers to be shown in the output, in addition to the accession and/or locus name.

Most preferably, sequence comparisons are conducted using the simple BLAST search algorithm provided at <http://www.ncbi.nlm.nih.gov/BLAST>. In some embodiments of the present invention, no gap penalties are used when determining sequence identity.

HYBRIDISATION

- 5 The present invention also encompasses nucleotide sequences that are capable of hybridising to the sequences presented herein, or any fragment or derivative thereof, or to the complement of any of the above.

Hybridization means a "process by which a strand of nucleic acid joins with a complementary strand through base pairing" (Coombs J (1994) Dictionary of Biotechnology, Stockton Press, New York NY) as well as the process of amplification as carried out in polymerase chain reaction technologies as described in Dieffenbach CW and GS Dveksler (1995, PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NY).

15 Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA), and confer a defined "stringency" as explained below.

Nucleotide sequences of the invention capable of selectively hybridising to the nucleotide sequences presented herein, or to their complement, will be generally at least 70%, 20 preferably at least 75%, more preferably at least 85 or 90% and even more preferably at least 95% or 98% homologous to the corresponding nucleotide sequences presented herein over a region of at least 20, preferably at least 25 or 30, for instance at least 40, 60 or 100 or more contiguous nucleotides. Preferred nucleotide sequences of the invention will comprise regions homologous to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 7 or 25 SEQ ID NO: 10, preferably at least 70%, 80% or 90% and more preferably at least 95% homologous to one of the sequences.

- The term "selectively hybridizable" means that the nucleotide sequence used as a probe is used under conditions where a target nucleotide sequence of the invention is found to hybridize to the probe at a level significantly above background. The background hybridization may occur because of other nucleotide sequences present, for example, in the
- 5 cDNA or genomic DNA library being screened. In this event, background implies a level of signal generated by interaction between the probe and a non-specific DNA member of the library which is less than 10 fold, preferably less than 100 fold as intense as the specific interaction observed with the target DNA. The intensity of interaction may be measured, for example, by radiolabelling the probe, e.g. with ^{32}P .
- 10 Also included within the scope of the present invention are nucleotide sequences that are capable of hybridizing to the nucleotide sequences presented herein under conditions of intermediate to maximal stringency. Hybridization conditions are based on the melting temperature (T_m) of the nucleic acid binding complex, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA), and confer a defined "stringency" as explained below.

Maximum stringency typically occurs at about $T_m - 5^\circ\text{C}$ (5°C below the T_m of the probe); high stringency at about 5°C to 10°C below T_m ; intermediate stringency at about 10°C to 20°C below T_m ; and low stringency at about 20°C to 25°C below T_m . As will be understood by those of skill in the art, a maximum stringency hybridization can be used to identify or detect identical nucleotide sequences while an intermediate (or low) stringency hybridization can be used to identify or detect similar or related nucleotide sequences.

In a preferred embodiment, the present invention covers nucleotide sequences that can hybridise to one or more of the Tarzan GPCR nucleotide sequences of the present invention under stringent conditions (e.g. 65°C and $0.1\times\text{SSC}$ { $1\times\text{SSC} = 0.15 \text{ M NaCl}, 0.015 \text{ M Na}_3\text{Citrate pH }7.0$ }). Where the nucleotide sequence of the invention is double-stranded, both strands of the duplex, either individually or in combination, are encompassed by the present invention. Where the nucleotide sequence is single-stranded, it is to be understood

that the complementary sequence of that nucleotide sequence is also included within the scope of the present invention.

The present invention also encompasses nucleotide sequences that are capable of hybridising to the sequences that are complementary to the sequences presented herein, or 5 any fragment or derivative thereof. Likewise, the present invention encompasses nucleotide sequences that are complementary to sequences that are capable of hybridising to the sequence of the present invention. These types of nucleotide sequences are examples of variant nucleotide sequences. In this respect, the term "variant" encompasses sequences that are complementary to sequences that are capable of hybridising to the 10 nucleotide sequences presented herein. Preferably, however, the term "variant" encompasses sequences that are complementary to sequences that are capable of hybridising under stringent conditions (eg. 65°C and 0.1xSSC {1xSSC = 0.15 M NaCl, 0.015 Na₃ citrate pH 7.0}) to the nucleotide sequences presented herein.

CLONING OF TARZAN GPCR AND HOMOLOGUES

15 The present invention also encompasses nucleotide sequences that are complementary to the sequences presented here, or any fragment or derivative thereof. If the sequence is complementary to a fragment thereof then that sequence can be used as a probe to identify and clone similar GPCR sequences in other organisms etc.

The present invention thus enables the cloning of Tarzan GPCR, its homologues 20 and other structurally or functionally related genes from human and other species such as mouse, pig, sheep, etc to be accomplished. Polynucleotides of the invention, which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 7 or SEQ ID NO: 10 or a fragment thereof, may be used as hybridization probes for cDNA and genomic DNA, to isolate partial or 25 full-length cDNAs and genomic clones encoding Tarzan GPCR from appropriate libraries. Such probes may also be used to isolate cDNA and genomic clones of other genes (including genes encoding homologues and orthologues from species other than human) that have sequence similarity, preferably high sequence similarity, to the Tarzan GPCR

gene. Hybridization screening, cloning and sequencing techniques are known to those of skill in the art and are described in, for example, Sambrook et al (*supra*).

Typically nucleotide sequences suitable for use as probes are 70% identical, preferably 80% identical, more preferably 90% identical, even more preferably 95% 5 identical to that of the referent. The probes generally will comprise at least 15 nucleotides. Preferably, such probes will have at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will range between 150 and 500 nucleotides, more particularly about 300 nucleotides.

In one embodiment, to obtain a polynucleotide encoding a Tarzan GPCR 10 polypeptide, including homologues and orthologues from species other than human, comprises the steps of screening an appropriate library under stringent hybridization conditions with a labelled probe having the SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 7 or SEQ ID NO: 10 or a fragment thereof and isolating partial or full-length cDNA and genomic clones containing said polynucleotide sequence. Such 15 hybridization techniques are well known to those of skill in the art. Stringent hybridization conditions are as defined above or alternatively conditions under overnight incubation at 42 degrees C. in a solution comprising: 50% formamide, 5XSSC (150 mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5XDenhardt's solution, 10% dextran sulphate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed 20 by washing the filters in 0.1XSSC at about 65 degrees C.

FUNCTIONAL ASSAY FOR TARZAN GPCR

The cloned putative Tarzan GPCR polynucleotides may be verified by sequence analysis or functional assays. For example, the putative Tarzan GPCR or homologue may be assayed for receptor activity as follows. Capped RNA transcripts from linearized 25 plasmid templates encoding the Tarzan receptor cDNAs of the invention are synthesized *in vitro* with RNA polymerases in accordance with standard procedures. *In vitro* transcripts are suspended in water at a final concentration of 0.2 mg/ml. Ovarian lobes are removed from adult female toads, Stage V defolliculated oocytes are obtained, and RNA

transcripts (10 ng/oocyte) are injected in a 50 nl bolus using a microinjection apparatus. Two electrode voltage clamps are used to measure the currents from individual *Xenopus* oocytes in response to agonist exposure. Recordings are made in Ca²⁺ free Barth's medium at room temperature. The *Xenopus* system may also be used to screen known 5 ligands and tissue/cell extracts for activating ligands, as described in further detail below.

EXPRESSION ASSAYS FOR TARZAN GPCR

In order to design useful therapeutics for treating Tarzan GPCR associated diseases, it is useful to determine the expression profile of Tarzan (whether wild-type or a particular mutant). Thus, methods known in the art may be used to determine the organs, 10 tissues and cell types (as well as the developmental stages) in which Tarzan is expressed. For example, traditional or "electronic" Northerns may be conducted. Reverse-transcriptase PCR (RT-PCR) may also be employed to assay expression of the Tarzan gene or mutant. More sensitive methods for determining the expression profile of Tarzan include RNase protection assays, as known in the art.

15 Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (Sambrook, supra, ch. 7 and Ausubel, F. M. et al. supra, ch. 4 and 16.) Analogous computer techniques ("electronic Northerns") applying BLAST may be used to search for 20 identical or related molecules in nucleotide databases such as GenBank or the LIFESEQ database (Incyte Pharmaceuticals). This type of analysis has advantages in that they may be faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

25 The polynucleotides and polypeptides of the present invention, including the probes described above, may be employed as research reagents and materials for discovery of treatments and diagnostics to animal and human disease, as explained in further detail elsewhere in this document.

EXPRESSION OF TARZAN GPCR POLYPEPTIDES

The invention includes a process for producing a Tarzan GPCR polypeptide. The method comprises in general culturing a host cell comprising a nucleic acid encoding Tarzan GPCR polypeptide, or a homologue, variant, or derivative thereof, under suitable 5 conditions (i.e., conditions in which the Tarzan GPCR polypeptide is expressed).

In order to express a biologically active Tarzan GPCR, the nucleotide sequences encoding Tarzan GPCR or homologues, variants, or derivatives thereof are inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

10 Methods which are well known to those skilled in the art are used to construct expression vectors containing sequences encoding Tarzan GPCR and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described in Sambrook, J. et al. (1989; Molecular Cloning, A
15 Laboratory Manual, ch. 4, 8, and 16-17, Cold Spring Harbor Press, Plainview, N.Y.) and Ausubel, F. M. et al. (1995 and periodic supplements; Current Protocols in Molecular Biology, ch. 9, 13, and 16, John Wiley & Sons, New York, N.Y.).

A variety of expression vector/host systems may be utilized to contain and express 20 sequences encoding Tarzan GPCR. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or 25 animal cell systems. The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions of the vector (i.e., enhancers, promoters, and 5' and 3' untranslated regions) which interact

with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems,
5 inducible promoters such as the hybrid lacZ promoter of the BLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSPORT1 plasmid (GIBCO/BRL), and the like, may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or
enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO, and
storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences)
10 may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding Tarzan GPCR, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending
15 upon the use intended for Tarzan GPCR. For example, when large quantities of Tarzan GPCR are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, multifunctional E. coli cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding Tarzan GPCR may be
20 ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β -galactosidase so that a hybrid protein is produced, pIN vectors (Van Heeke, G. and S. M. Schuster (1989) J. Biol. Chem. 264:5503-5509), and the like. pGEX vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins
25 are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used. For reviews, see Ausubel (supra) and Grant et al. (1987; Methods Enzymol. 153:516-544).

In cases where plant expression vectors are used, the expression of sequences encoding Tarzan GPCR may be driven by any of a number of promoters. For example, 5 viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) EMBO J. 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, 10 R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a 15 number of generally available reviews. (See, for example, Hobbs, S. or Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp. 191-196.).

An insect system may also be used to express Tarzan GPCR. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequences encoding Tarzan GPCR may be cloned into a non-essential region of the 20 virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of Tarzan GPCR will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which Tarzan GPCR may be expressed. (Engelhard, E. K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227.)

25 In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding Tarzan GPCR may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable

of expressing Tarzan GPCR in infected host cells. (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

5 Thus, for example, the Tarzan receptors of the present invention are expressed in either human embryonic kidney 293 (HEK293) cells or adherent dhfr CHO cells. To maximize receptor expression, typically all 5' and 3' untranslated regions (UTRs) are removed from the receptor cDNA prior to insertion into a pCDN or pCDNA3 vector. The cells are transfected with individual receptor cDNAs by lipofectin and selected in the
10 presence of 400 mg/ml G418. After 3 weeks of selection, individual clones are picked and expanded for further analysis. HEK293 or CHO cells transfected with the vector alone serve as negative controls. To isolate cell lines stably expressing the individual receptors, about 24 clones are typically selected and analyzed by Northern blot analysis. Receptor mRNAs are generally detectable in about 50% of the G418-resistant clones analyzed.

15 Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

20 Specific initiation signals may also be used to achieve more efficient translation of sequences encoding Tarzan GPCR. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding Tarzan GPCR and its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous
25 translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular cell system used,

such as those described in the literature. (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such
5 modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "pro" form of the protein may also be used to facilitate correct insertion, folding, and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO,
10 HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Bethesda, Md.) and may be chosen to ensure the correct modification and processing of the foreign protein.

For long term, high yield production of recombinant proteins, stable expression is preferred. For example, cell lines capable of stably expressing Tarzan GPCR can be
15 transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth
20 and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase genes
25 (Wigler, M. et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase genes (Lowy, I. et al. (1980) Cell 22:817-23), which can be employed in tk⁻ or apr⁻ cells, respectively. Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, dhfr confers resistance to methotrexate (Wigler, M. et al.
(1980) Proc. Natl. Acad. Sci. 77:3567-70); npt confers resistance to the aminoglycosides

neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14); and als or pat confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine. (Hartman, S. C. and R. C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-51.) Recently, the use of visible markers has gained popularity with such markers as anthocyanins, β -glucuronidase and its substrate GUS, and luciferase and its substrate luciferin. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (Rhodes, C. A. et al. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding Tarzan GPCR is inserted within a marker gene sequence, transformed cells containing sequences encoding Tarzan GPCR can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding Tarzan GPCR under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding Tarzan GPCR and express Tarzan GPCR may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA--DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

The presence of polynucleotide sequences encoding Tarzan GPCR can be detected by DNA--DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding Tarzan GPCR. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding Tarzan GPCR to detect transformants containing DNA or RNA encoding Tarzan GPCR.

A variety of protocols for detecting and measuring the expression of Tarzan GPCR, using either polyclonal or monoclonal antibodies specific for the protein, are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on Tarzan GPCR is preferred, but a competitive binding assay may be employed. These and other assays are well described in the art, for example, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, Section IV, APS Press, St Paul, Minn.) and in Maddox, D. E. et al. (1983; J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding Tarzan GPCR include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding Tarzan GPCR, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Pharmacia & Upjohn (Kalamazoo, Mich.), Promega (Madison, Wis.), and U.S. Biochemical Corp. (Cleveland, Ohio). Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding Tarzan GPCR may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be located in the cell membrane, secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing

- polynucleotides which encode Tarzan GPCR may be designed to contain signal sequences which direct secretion of Tarzan GPCR through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding Tarzan GPCR to nucleotide sequences encoding a polypeptide domain which will facilitate purification of
- 5 soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences,
- 10 such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, Calif.), between the purification domain and the Tarzan GPCR encoding sequence may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing Tarzan GPCR and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate
- 15 purification on immobilized metal ion affinity chromatography (IMIAC; described in Porath, J. et al. (1992) Prot. Exp. Purif. 3: 263-281), while the enterokinase cleavage site provides a means for purifying Tarzan GPCR from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D. J. et al. (1993; DNA Cell Biol. 12:441-453).
- 20 Fragments of Tarzan GPCR may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (Merrifield J. (1963) J. Am. Chem. Soc. 85:2149-2154.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the Applied Biosystems 431A peptide synthesizer (Perkin Elmer). Various fragments of
- 25 Tarzan GPCR may be synthesized separately and then combined to produce the full length molecule.

BIOSENSORS

The Tarzan polypeptides, nucleic acids, probes, antibodies, expression vectors and ligands are useful as (and for the production of) biosensors.

According to Aizawa (1988), *Anal. Chem. Symp.* 17: 683, a biosensor is defined as being a unique combination of a receptor for molecular recognition, for example a selective layer with immobilized antibodies or receptors such as a Tarzan G-protein coupled receptor, and a transducer for transmitting the values measured. One group of 5 such biosensors will detect the change which is caused in the optical properties of a surface layer due to the interaction of the receptor with the surrounding medium. Among such techniques may be mentioned especially ellipso-metry and surface plasmon resonance. Biosensors incorporating Tarzan may be used to detect the presence or level of Tarzan ligands, for example, nucleotides such as purines or purine analogues, or analogues 10 of these ligands. The construction of such biosensors is well known in the art.

Thus, cell lines expressing Tarzan receptor may be used as reporter systems for detection of ligands such as ATP via receptor-promoted formation of [3H]inositol phosphates or other second messengers (Watt et al., 1998, *J Biol Chem* May 29;273(22):14053-8). Receptor-ligand biosensors are also described in Hoffman et al., 15 2000, *Proc Natl Acad Sci USA* Oct 10;97(21):11215-20. Optical and other biosensors comprising Tarzan may also be used to detect the level or presence of interaction with G-proteins and other proteins, as described by, for example, Figler et al, 1997, *Biochemistry* Dec 23;36(51):16288-99 and Sarrio et al., 2000, *Mol Cell Biol* 2000 Jul;20(14):5164-74). Sensor units for biosensors are described in, for example, US 5,492,840.

20 SCREENER ASSAYS

The Tarzan GPCR polypeptide of the present invention, including homologues, variants, and derivatives, whether natural or recombinant, may be employed in a screening process for compounds which bind the receptor and which activate (agonists) or inhibit activation of (antagonists) of Tarzan. Thus, polypeptides of the invention may also be used 25 to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics. See Coligan et al., *Current Protocols in Immunology* 1(2):Chapter 5 (1991).

- Tarzan GPCR polypeptides are responsible for many biological functions, including many pathologies. Accordingly, it is desirous to find compounds and drugs which stimulate Tarzan GPCR on the one hand and which can inhibit the function of Tarzan GPCR on the other hand. In general, agonists and antagonists are employed for therapeutic and prophylactic purposes for such conditions as infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; diabetes; obesity; anorexia; bulimia; asthma; parkinson's disease; thrombosis; acute heart failure; hypotension; hypertension; erectile dysfunction; urinary retention; metabolic bone diseases such as osteoporosis and osteo petrosis; angina pectoris; myocardial infarction; ulcers; asthma; allergies; rheumatoid arthritis; inflammatory bowel disease; irritable bowel syndrome benign prostatic hypertrophy; and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles dela Tourett's syndrome.
- Rational design of candidate compounds likely to be able to interact with Tarzan GPCR protein may be based upon structural studies of the molecular shapes of a polypeptide according to the invention. One means for determining which sites interact with specific other proteins is a physical structure determination, e.g., X-ray crystallography or two-dimensional NMR techniques. These will provide guidance as to which amino acid residues form molecular contact regions. For a detailed description of protein structural determination, see, e.g., Blundell and Johnson (1976) *Protein Crystallography*, Academic Press, New York.

An alternative to rational design uses a screening procedure which involves in general producing appropriate cells which express the Tarzan receptor polypeptide of the present invention on the surface thereof. Such cells include cells from animals, yeast, *Drosophila* or *E. coli*. Cells expressing the receptor (or cell membrane containing the expressed receptor) are then contacted with a test compound to observe binding, or stimulation or inhibition of a functional response. For example, *Xenopus* oocytes may be injected with Tarzan mRNA or polypeptide, and currents induced by exposure to test

compounds measured by use of voltage clamps measured, as described in further detail elsewhere.

Furthermore, microphysiometric assays may be employed to assay Tarzan receptor activity. Activation of a wide variety of secondary messenger systems results in extrusion 5 of small amounts of acid from a cell. The acid formed is largely as a result of the increased metabolic activity required to fuel the intracellular signalling process. The pH changes in the media surrounding the cell are very small but are detectable by, for example, the CYTOSENSOR microphysiometer (Molecular Devices Ltd., Menlo Park, Calif.). The CYTOSENSOR is thus capable of detecting the activation of a receptor which is coupled 10 to an energy utilizing intracellular signaling pathway such as the G-protein coupled receptor of the present invention.

Instead of testing each candidate compound individually with the Tarzan receptor, a library or bank of candidate ligands may advantageously be produced and screened. Thus, for example, a bank of over 200 putative receptor ligands has been assembled for 15 screening. The bank comprises: transmitters, hormones and chemokines known to act via a human seven transmembrane (7TM) receptor; naturally occurring compounds which may be putative agonists for a human 7TM receptor, non-mammalian, biologically active peptides for which a mammalian counterpart has not yet been identified; and compounds not found in nature, but which activate 7TM receptors with unknown natural ligands. This 20 bank is used to screen the receptor for known ligands, using both functional (i.e. calcium, cAMP, microphysiometer, oocyte electrophysiology, etc, see elsewhere) as well as binding assays as described in further detail elsewhere. However, a large number of mammalian receptors exist for which there remains, as yet, no cognate activating ligand (agonist) or deactivating ligand (antagonist). Thus, active ligands for these receptors may 25 not be included within the ligands banks as identified to date. Accordingly, the Tarzan receptor of the invention is also functionally screened (using calcium, cAMP, microphysiometer, oocyte electrophysiology, etc., functional screens) against tissue extracts to identify natural ligands. Extracts that produce positive functional responses can be sequentially subfractionated, with the fractions being assayed as described here, until 30 an activating ligand is isolated and identified.

7TM receptors which are expressed in HEK 293 cells have been shown to be coupled functionally to activation of PLC and calcium mobilization and/or cAMP stimulation or inhibition. One screening technique therefore includes the use of cells which express the Tarzan GPCR receptor of this invention (for example, transfected *Xenopus* oocytes, CHO or HEK293 cells) in a system which measures extracellular pH or intracellular calcium changes caused by receptor activation. In this technique, compounds may be contacted with cells expressing the receptor polypeptide of the present invention. A second messenger response, e.g., signal transduction, pH changes, or changes in calcium level, is then measured to determine whether the potential compound activates or inhibits the receptor.

In such experiments, basal calcium levels in the HEK 293 cells in receptor-transfected or vector control cells are observed to be in the normal, 100 nM to 200 nM, range. HEK 293 cells expressing Tarzan GPCR or recombinant Tarzan GPCR are loaded with fura 2 and in a single day more than 150 selected ligands or tissue/cell extracts are evaluated for agonist induced calcium mobilization. Similarly, HEK 293 cells expressing Tarzan GPCR or recombinant Tarzan GPCR are evaluated for the stimulation or inhibition of cAMP production using standard cAMP quantitation assays. Agonists presenting a calcium transient or cAMP fluctuation are tested in vector control cells to determine if the response is unique to the transfected cells expressing receptor.

Another method involves screening for receptor inhibitors by determining inhibition or stimulation of Tarzan receptor-mediated cAMP and/or adenylate cyclase accumulation. Such a method involves transfecting a eukaryotic cell with the receptor of this invention to express the receptor on the cell surface. The cell is then exposed to potential antagonists in the presence of the receptor of this invention. The amount of cAMP accumulation is then measured. If the potential antagonist binds the receptor, and thus inhibits receptor binding, the levels of receptor-mediated cAMP, or adenylate cyclase, activity will be reduced or increased.

Another method for detecting agonists or antagonists for the receptor of the present invention is the yeast based technology as described in U.S. Pat. No. 5,482,835, incorporated by reference herein.

Where the candidate compounds are proteins, in particular antibodies or peptides, 5 libraries of candidate compounds may be screened using phage display techniques. Phage display is a protocol of molecular screening which utilises recombinant bacteriophage. The technology involves transforming bacteriophage with a gene that encodes one compound from the library of candidate compounds, such that each phage or phagemid expresses a particular candidate compound. The transformed bacteriophage (which 10 preferably is tethered to a solid support) expresses the appropriate candidate compound and displays it on their phage coat. Specific candidate compounds which are capable of binding to a polypeptide or peptide of the invention are enriched by selection strategies based on affinity interaction. The successful candidate agents are then characterised. Phage display has advantages over standard affinity ligand screening technologies. The 15 phage surface displays the candidate agent in a three dimensional configuration, more closely resembling its naturally occurring conformation. This allows for more specific and higher affinity binding for screening purposes.

Another method of screening a library of compounds utilises eukaryotic or prokaryotic host cells which are stably transformed with recombinant DNA molecules 20 expressing a library of compounds. Such cells, either in viable or fixed form, can be used for standard binding-partner assays. See also Parce *et al.* (1989) Science 246:243-247; and Owicki *et al.* (1990) Proc. Nat'l Acad. Sci. USA 87:4007-4011, which describe sensitive methods to detect cellular responses. Competitive assays are particularly useful, where the 25 cells expressing the library of compounds are contacted or incubated with a labelled antibody known to bind to a Tarzan polypeptide of the present invention, such as ¹²⁵I- antibody, and a test sample such as a candidate compound whose binding affinity to the binding composition is being measured. The bound and free labelled binding partners for the polypeptide are then separated to assess the degree of binding. The amount of test sample bound is inversely proportional to the amount of labelled antibody binding to the 30 polypeptide.

Any one of numerous techniques can be used to separate bound from free binding partners to assess the degree of binding. This separation step could typically involve a procedure such as adhesion to filters followed by washing, adhesion to plastic following by washing, or centrifugation of the cell membranes.

5 Still another approach is to use solubilized, unpurified or solubilized purified polypeptide or peptides, for example extracted from transformed eukaryotic or prokaryotic host cells. This allows for a "molecular" binding assay with the advantages of increased specificity, the ability to automate, and high drug test throughput.

Another technique for candidate compound screening involves an approach which
10 provides high throughput screening for new compounds having suitable binding affinity, e.g., to a polypeptide of the invention, and is described in detail in International Patent application no. WO 84/03564 (Commonwealth Serum Labs.), published on September 13 1984. First, large numbers of different small peptide test compounds are synthesized on a solid substrate, e.g., plastic pins or some other appropriate surface; see Fodor *et al.* (1991).
15 Then all the pins are reacted with solubilized polypeptide of the invention and washed. The next step involves detecting bound polypeptide. Compounds which interact specifically with the polypeptide will thus be identified.

Ligand binding assays provide a direct method for ascertaining receptor pharmacology and are adaptable to a high throughput format. The purified ligand for a
20 receptor may be radiolabeled to high specific activity (50-2000 Ci/mmol) for binding studies. A determination is then made that the process of radiolabeling does not diminish the activity of the ligand towards its receptor. Assay conditions for buffers, ions, pH and other modulators such as nucleotides are optimized to establish a workable signal to noise ratio for both membrane and whole cell receptor sources. For these assays, specific
25 receptor binding is defined as total associated radioactivity minus the radioactivity measured in the presence of an excess of unlabeled competing ligand. Where possible, more than one competing ligand is used to define residual nonspecific binding.

The assays may simply test binding of a candidate compound wherein adherence to the cells bearing the receptor is detected by means of a label directly or indirectly associated with the candidate compound or in an assay involving competition with a labeled competitor. Further, these assays may test whether the candidate compound results 5 in a signal generated by activation of the receptor, using detection systems appropriate to the cells bearing the receptor at their surfaces. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed.

Further, the assays may simply comprise the steps of mixing a candidate 10 compound with a solution containing a Tarzan GPCR polypeptide to form a mixture, measuring Tarzan GPCR activity in the mixture, and comparing the Tarzan GPCR activity of the mixture to a standard.

The Tarzan GPCR cDNA, protein and antibodies to the protein may also be used to configure assays for detecting the effect of added compounds on the production of Tarzan 15 GPCR mRNA and protein in cells. For example, an ELISA may be constructed for measuring secreted or cell associated levels of Tarzan GPCR protein using monoclonal and polyclonal antibodies by standard methods known in the art, and this can be used to discover agents which may inhibit or enhance the production of Tarzan GPCR (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues. Standard 20 methods for conducting screening assays are well understood in the art.

Examples of potential Tarzan GPCR antagonists include antibodies or, in some cases, nucleotides and their analogues, including purines and purine analogues, oligonucleotides or proteins which are closely related to the ligand of the Tarzan GPCR, e.g., a fragment of the ligand, or small molecules which bind to the receptor but do not 25 elicit a response, so that the activity of the receptor is prevented.

The present invention therefore also provides a compound capable of binding specifically to a Tarzan polypeptide and/or peptide of the present invention.

The term "compound" refers to a chemical compound (naturally occurring or synthesised), such as a biological macromolecule (e.g., nucleic acid, protein, non-peptide, or organic molecule), or an extract made from biological materials such as bacteria, plants, fungi, or animal (particularly mammalian) cells or tissues, or even an inorganic element or
5 molecule. Preferably the compound is an antibody.

The materials necessary for such screening to be conducted may be packaged into a screening kit. Such a screening kit is useful for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, etc. for Tarzan GPCR polypeptides or compounds which decrease or enhance the production of Tarzan GPCR polypeptides. The screening kit
10 comprises: (a) a Tarzan GPCR polypeptide; (b) a recombinant cell expressing a Tarzan GPCR polypeptide; (c) a cell membrane expressing a Tarzan GPCR polypeptide; or (d) antibody to a Tarzan GPCR polypeptide. The screening kit may optionally comprise instructions for use.

TRANSGENIC ANIMALS

15 The present invention further encompasses transgenic animals capable of expressing natural or recombinant Tarzan GPCR, or a homologue, variant or derivative, at elevated or reduced levels compared to the normal expression level. Included are transgenic animals ("Tarzan knockout"s) which do not express functional Tarzan receptor. The Tarzan knockouts may arise as a result of functional disruption of the Tarzan gene or
20 any portion of that gene, including one or more loss of function mutations, including a deletion or replacement, of the Tarzan gene. The mutations include single point mutations, and may target coding or non-coding regions of Tarzan.

25 Preferably, such a transgenic animal is a non-human mammal, such as a pig, a sheep or a rodent. Most preferably the transgenic animal is a mouse or a rat. Such transgenic animals may be used in screening procedures to identify agonists and/or antagonists of Tarzan GPCR, as well as to test for their efficacy as treatments for diseases *in vivo*.

For example, transgenic animals that have been engineered to be deficient in the production of Tarzan GPCR may be used in assays to identify agonists and/or antagonists of Tarzan GPCR. One assay is designed to evaluate a potential drug (aa candidate ligand or compound) to determine if it produces a physiological response in the absence of

5 Tarzan GPCR receptors. This may be accomplished by administering the drug to a transgenic animal as discussed above, and then assaying the animal for a particular response. Any physiological parameter could be measured in this assay.

Tissues derived from the Tarzan knockout animals may be used in receptor binding assays to determine whether the potential drug (a candidate ligand or compound) binds to

10 the Tarzan receptor. Such assays can be conducted by obtaining a first receptor preparation from the transgenic animal engineered to be deficient in Tarzan receptor production and a second receptor preparation from a source known to bind any identified Tarzan ligands or compounds. In general, the first and second receptor preparations will be similar in all respects except for the source from which they are obtained. For example, if

15 brain tissue from a transgenic animal (such as described above and below) is used in an assay, comparable brain tissue from a normal (wild type) animal is used as the source of the second receptor preparation. Each of the receptor preparations is incubated with a ligand known to bind to Tarzan receptors, both alone and in the presence of the candidate ligand or compound. Preferably, the candidate ligand or compound will be examined at

20 several different concentrations.

The extent to which binding by the known ligand is displaced by the test compound is determined for both the first and second receptor preparations. Tissues derived from transgenic animals may be used in assays directly or the tissues may be processed to isolate membranes or membrane proteins, which are themselves used in the assays. A preferred transgenic animal is the mouse. The ligand may be labeled using any means compatible with binding assays. This would include, without limitation, radioactive, enzymatic, fluorescent or chemiluminescent labeling (as well as other labelling techniques as described in further detail above).

Furthermore, antagonists of Tarzan GPCR receptor may be identified by administering candidate compounds, etc, to wild type animals expressing functional Tarzan, and animals identified which exhibit any of the phenotypic characteristics associated with reduced or abolished expression of Tarzan receptor function.

5 Detailed methods for generating non-human transgenic animal are described in further detail below. Transgenic gene constructs can be introduced into the germ line of an animal to make a transgenic mammal. For example, one or several copies of the construct may be incorporated into the genome of a mammalian embryo by standard transgenic techniques.

10 In an exemplary embodiment, the transgenic non-human animals of the invention are produced by introducing transgenes into the germline of the non-human animal. Embryonal target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the stage of development of the embryonal target cell. The specific line(s) of any animal used to practice this invention are
15 selected for general good health, good embryo yields, good pronuclear visibility in the embryo, and good reproductive fitness. In addition, the haplotype is a significant factor.

Introduction of the transgene into the embryo can be accomplished by any means known in the art such as, for example, microinjection, electroporation, or lipofection. For example, the Tarzan receptor transgene can be introduced into a mammal by
20 microinjection of the construct into the pronuclei of the fertilized mammalian egg(s) to cause one or more copies of the construct to be retained in the cells of the developing mammal(s). Following introduction of the transgene construct into the fertilized egg, the egg may be incubated *in vitro* for varying amounts of time, or reimplanted into the surrogate host, or both. *In vitro* incubation to maturity is within the scope of this invention.
25 One common method is to incubate the embryos *in vitro* for about 1-7 days, depending on the species, and then reimplant them into the surrogate host.

The progeny of the transgenically manipulated embryos can be tested for the presence of the construct by Southern blot analysis of the segment of tissue. If one or more

copies of the exogenous cloned construct remains stably integrated into the genome of such transgenic embryos, it is possible to establish permanent transgenic mammal lines carrying the transgenically added construct.

The litters of transgenically altered mammals can be assayed after birth for the 5 incorporation of the construct into the genome of the offspring. Preferably, this assay is accomplished by hybridizing a probe corresponding to the DNA sequence coding for the desired recombinant protein product or a segment thereof onto chromosomal material from the progeny. Those mammalian progeny found to contain at least one copy of the construct in their genome are grown to maturity.

10 For the purposes of this invention a zygote is essentially the formation of a diploid cell which is capable of developing into a complete organism. Generally, the zygote will be comprised of an egg containing a nucleus formed, either naturally or artificially, by the fusion of two haploid nuclei from a gamete or gametes. Thus, the gamete nuclei must be ones which are naturally compatible, i.e., ones which result in a viable zygote capable of 15 undergoing differentiation and developing into a functioning organism. Generally, a euploid zygote is preferred. If an aneuploid zygote is obtained, then the number of chromosomes should not vary by more than one with respect to the euploid number of the organism from which either gamete originated.

In addition to similar biological considerations, physical ones also govern the 20 amount (e.g., volume) of exogenous genetic material which can be added to the nucleus of the zygote or to the genetic material which forms a part of the zygote nucleus. If no genetic material is removed, then the amount of exogenous genetic material which can be added is limited by the amount which will be absorbed without being physically disruptive. Generally, the volume of exogenous genetic material inserted will not exceed 25 about 10 picoliters. The physical effects of addition must not be so great as to physically destroy the viability of the zygote. The biological limit of the number and variety of DNA sequences will vary depending upon the particular zygote and functions of the exogenous genetic material and will be readily apparent to one skilled in the art, because the genetic material, including the exogenous genetic material, of the resulting zygote must be

biologically capable of initiating and maintaining the differentiation and development of the zygote into a functional organism.

The number of copies of the transgene constructs which are added to the zygote is dependent upon the total amount of exogenous genetic material added and will be the 5 amount which enables the genetic transformation to occur. Theoretically only one copy is required; however, generally, numerous copies are utilized, for example, 1,000-20,000 copies of the transgene construct, in order to insure that one copy is functional. As regards the present invention, there will often be an advantage to having more than one functioning copy of each of the inserted exogenous DNA sequences to enhance the 10 phenotypic expression of the exogenous DNA sequences.

Any technique which allows for the addition of the exogenous genetic material into nucleic genetic material can be utilized so long as it is not destructive to the cell, nuclear membrane or other existing cellular or genetic structures. The exogenous genetic material is preferentially inserted into the nucleic genetic material by microinjection.

15 Microinjection of cells and cellular structures is known and is used in the art.

Reimplantation is accomplished using standard methods. Usually, the surrogate host is anesthetized, and the embryos are inserted into the oviduct. The number of embryos implanted into a particular host will vary by species, but will usually be comparable to the number of off spring the species naturally produces.

20 Transgenic offspring of the surrogate host may be screened for the presence and/or expression of the transgene by any suitable method. Screening is often accomplished by Southern blot or Northern blot analysis, using a probe that is complementary to at least a portion of the transgene. Western blot analysis using an antibody against the protein encoded by the transgene may be employed as an alternative or additional method for 25 screening for the presence of the transgene product. Typically, DNA is prepared from tail tissue and analyzed by Southern analysis or PCR for the transgene. Alternatively, the tissues or cells believed to express the transgene at the highest levels are tested for the

presence and expression of the transgene using Southern analysis or PCR, although any tissues or cell types may be used for this analysis.

- Alternative or additional methods for evaluating the presence of the transgene include, without limitation, suitable biochemical assays such as enzyme and/or
- 5 immunological assays, histological stains for particular marker or enzyme activities, flow cytometric analysis, and the like. Analysis of the blood may also be useful to detect the presence of the transgene product in the blood, as well as to evaluate the effect of the transgene on the levels of various types of blood cells and other blood constituents.

- Progeny of the transgenic animals may be obtained by mating the transgenic
- 10 animal with a suitable partner, or by in vitro fertilization of eggs and/or sperm obtained from the transgenic animal. Where mating with a partner is to be performed, the partner may or may not be transgenic and/or a knockout; where it is transgenic, it may contain the same or a different transgene, or both. Alternatively, the partner may be a parental line. Where in vitro fertilization is used, the fertilized embryo may be implanted into a
- 15 surrogate host or incubated in vitro, or both. Using either method, the progeny may be evaluated for the presence of the transgene using methods described above, or other appropriate methods.

- The transgenic animals produced in accordance with the present invention will include exogenous genetic material. As set out above, the exogenous genetic material will,
- 20 in certain embodiments, be a DNA sequence which results in the production of a Tarzan GPCR receptor. Further, in such embodiments the sequence will be attached to a transcriptional control element, e.g., a promoter, which preferably allows the expression of the transgene product in a specific type of cell.

- Retroviral infection can also be used to introduce transgene into a non-human
- 25 animal. The developing non-human embryo can be cultured in vitro to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Jaenich, R. (1976) PNAS 73:1260-1264). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Manipulating the Mouse Embryo, Hogan eds.

(Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1986). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner et al. (1985) PNAS 82:6927-6931; Van der Putten et al. (1985) PNAS 82:6148-6152). Transfection is easily and efficiently obtained by culturing 5 the blastomeres on a monolayer of virus-producing cells (Van der Putten, *supra*; Stewart et al. (1987) EMBO J. 6:383-388). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahner et al. (1982) Nature 298:623-628). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells which formed the transgenic non-human 10 animal. Further, the founder may contain various retroviral insertions of the transgene at different positions in the genome which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line by intrauterine retroviral infection of the midgestation embryo (Jahner et al. (1982) *supra*).

A third type of target cell for transgene introduction is the embryonal stem cell 15 (ES). ES cells are obtained from pre-implantation embryos cultured in vitro and fused with embryos (Evans et al. (1981) Nature 292:154-156; Bradley et al. (1984) Nature 309:255-258; Gossler et al. (1986) PNAS 83: 9065-9069; and Robertson et al. (1986) Nature 322:445-448). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retrovirus-mediated transduction. Such transformed ES cells can 20 thereafter be combined with blastocysts from a non-human animal. The ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. For review see Jaenisch, R. (1988) Science 240:1468-1474.

We also provide non-human transgenic animals, where the transgenic animal is characterized by having an altered Tarzan gene, preferably as described above, as models 25 for Tarzan receptor function. Alterations to the gene include deletions or other loss of function mutations, introduction of an exogenous gene having a nucleotide sequence with targeted or random mutations, introduction of an exogenous gene from another species, or a combination thereof. The transgenic animals may be either homozygous or heterozygous for the alteration. The animals and cells derived therefrom are useful for screening 30 biologically active agents that may modulate Tarzanreceptor function. The screening

methods are of particular use for determining the specificity and action of potential therapies for infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; diabetes, obesity; anorexia; bulimia; asthma; parkinson's disease; thrombosis; acute heart failure; 5 hypotension; hypertension; erectile dysfunction; urinary retention; metabolic bone diseases such as osteoporosis and osteo petrosis; angina pectoris; myocardial infarction; ulcers; asthma; allergies; rheumatoid arthritis; inflammatory bowel disease; irritable bowel syndrome benign prostatic hypertrophy; and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental 10 retardation and dyskinesias, such as Huntington's disease or Gilles dela Tourett's syndrome. The animals are useful as a model to investigate the role of Tarzan receptors in normal brain, heart, spleen and liver function.

Another aspect of the invention pertains to a transgenic nonhuman animal having a functionally disrupted endogenous Tarzan gene but which also carries in its genome, and 15 expresses, a transgene encoding a heterologous Tarzan protein (i.e., a Tarzan from another species). Preferably, the animal is a mouse and the heterologous Tarzan is a human Tarzan. An animal, or cell lines derived from such an animal of the invention, which has been reconstituted with human Tarzan, can be used to identify agents that inhibit human Tarzan *in vivo* and *in vitro*. For example, a stimulus that induces signalling through human 20 Tarzan can be administered to the animal, or cell line, in the presence and absence of an agent to be tested and the response in the animal, or cell line, can be measured. An agent that inhibits human Tarzan *in vivo* or *in vitro* can be identified based upon a decreased response in the presence of the agent compared to the response in the absence of the agent.

The present invention also provides for a Tarzan GPCR deficient transgenic non- 25 human animal (a "Tarzan GPCR knock-out"). Such an animal is one which expresses lowered or no Tarzan GPCR activity, preferably as a result of an endogenous Tarzan GPCR genomic sequence being disrupted or deleted. The endogenous Tarzan GPCR genomic sequence may be replaced by a null allele, which may comprise non-functional portions of the wild-type Tarzan sequence. For example, the endogenous Tarzan GPCR 30 genomic sequence may be replaced by an allele of Tarzan comprising a disrupting

sequence which may comprise heterologous sequences, for example, reporter sequences and/or selectable markers. Preferably, the endogenous Tarzan GPCR genomic sequence in a Tarzan knock-out mouse is replaced by an allele of Tarzan in which one or more, preferably all, of the transmembrane sequences is replaced by such a disrupting sequence,

5 preferably a lacZ sequence and a neomycin resistance sequence. Preferably, the genomic Tarzan sequence which is functionally disrupted comprises a mouse Tarzan genomic sequence shown in SEQ ID NO: 10.

Preferably, such an animal expresses no GPCR activity. More preferably, the animal expresses no activity of the Tarzan GPCR shown as SEQ ID NO: 3, SEQ ID NO: 10, SEQ ID NO: 5 or SEQ ID NO: 8. Tarzan GPCR knock-outs may be generated by various means known in the art, as described in further detail below. A specific description of the construction of a Tarzan knock-out mouse is disclosed in Examples 1 *et seq.*

The present invention also pertains to a nucleic acid construct for functionally disrupting a Tarzan gene in a host cell. The nucleic acid construct comprises: a) a non-homologous replacement portion; b) a first homology region located upstream of the non-homologous replacement portion, the first homology region having a nucleotide sequence with substantial identity to a first Tarzan gene sequence; and c) a second homology region located downstream of the non-homologous replacement portion, the second homology region having a nucleotide sequence with substantial identity to a second Tarzan gene sequence, the second Tarzan gene sequence having a location downstream of the first Tarzan gene sequence in a naturally occurring endogenous Tarzan gene. Additionally, the first and second homology regions are of sufficient length for homologous recombination between the nucleic acid construct and an endogenous Tarzan gene in a host cell when the nucleic acid molecule is introduced into the host cell. In a preferred embodiment, the non-homologous replacement portion comprises an expression reporter, preferably including lacZ and a positive selection expression cassette, preferably including a neomycin phosphotransferase gene operatively linked to a regulatory element(s).

Preferably, the first and second Tarzan gene sequences are derived from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 7 or SEQ ID NO: 10, or a homologue, variant or derivative thereof.

Another aspect of the invention pertains to recombinant vectors into which the nucleic acid construct of the invention has been incorporated. Yet another aspect of the invention pertains to host cells into which the nucleic acid construct of the invention has been introduced to thereby allow homologous recombination between the nucleic acid construct and an endogenous Tarzan gene of the host cell, resulting in functional disruption of the endogenous Tarzan gene. The host cell can be a mammalian cell that normally expresses Tarzan from the liver, brain, spleen or heart, or a pluripotent cell, such as a mouse embryonic stem cell. Further development of an embryonic stem cell into which the nucleic acid construct has been introduced and homologously recombined with the endogenous Tarzan gene produces a transgenic nonhuman animal having cells that are descendant from the embryonic stem cell and thus carry the Tarzan gene disruption in their genome. Animals that carry the Tarzan gene disruption in their germline can then be selected and bred to produce animals having the Tarzan gene disruption in all somatic and germ cells. Such mice can then be bred to homozygosity for the Tarzan gene disruption.

A TARZAN GPCR deficient transgenic animal may be generated as follows:

CONSTRUCTION OF TARZAN GENE TARGETING VECTOR

Murine TARZAN genomic clones are isolated from a mouse large insert PAC library obtained from HGMP (Hinxton, UK) using a probe sequence amplified from a part of the predicted murine open reading frame cDNA sequence (SEQ ID NO: 4), using standard techniques. The isolated murine TARZAN genomic clones are then restriction mapped in the region of the TARZAN gene using small oligonucleotide probes and standard techniques.

The murine genomic locus is partially sequenced to enable the design of homologous arms to clone into the targeting vector. Two regions of DNA, typically

between 1 and 5kb in size, from either side of the region of the open reading frame to be deleted, called the 5' and 3' homology arms, are amplified by PCR and the fragments are cloned into the targeting vector. The position of these arms is chosen so that a homologous recombination event will functionally disrupt the TARZAN gene by deleting at least the 5 seven trans-membrane spanning regions. A targeting vector is prepared where the deleted TARZAN sequence is replaced with non-homologous sequences composed of an endogenous gene expression reporter (a frame-independent lacZ gene) upstream of a selection cassette composed of a promoted neomycin phosphotransferase (neo) gene, arranged in the same orientation as the TARZAN gene.

10 The mouse genomic region used in the knockout strategy for the Tarzan gene is depicted as SEQ ID NO: 10

Transfection and Analysis of Embryonal Stem Cells

15 Embryonal stem cells (Evans and Kaufman, 1981) are cultured on a neomycin resistant embryonal fibroblast feeder layer grown in Dulbecco's Modified Eagles medium supplemented with 20% Fetal Calf Serum, 10% new-born calf serum, 2 mM glutamine, non-essential amino acids, 100 µM 2-mercaptoethanol and 500 u/ml leukemia inhibitory factor. Medium is changed daily and ES cells are subcultured every three days. 5×10^6 ES cells are transfected with 5 µg of linearized plasmid by electroporation (25 µF capacitance and 400 Volts). 24 hours following electroporation the transfected cells are cultured for 9 days in medium containing 200 µg/ml neomycin. Clones are picked into 96 well plates, replicated and expanded before being screened by PCR to identify clones in which homologous recombination had occurred between the endogenous Tarzan gene and the targeting construct. Positive clones are typically identified at a rate of 1 to 5%. These clones were expanded to allow replicas to be frozen and sufficient high quality DNA to 20 be prepared for Southern blot confirmation of the targeting event using external 5' and 3' probes, all using standard procedures (Russ et al, 2000, *Nature* 2000 Mar 2; 404(6773):95-25 9).

Generation of Tarzan GPCR Deficient Mice

C57BL/6 female and male mice are mated and blastocysts are isolated at 3.5 days of gestation. 10-12 cells from a chosen clone are injected per blastocyst and 7-8 blastocysts are implanted in the uterus of a pseudopregnant F1 female.

- 5 A litter of chimeric pups are born containing several high level (up to 100%) agouti males (the agouti coat colour indicates the contribution of cells descended from the targeted clone). The male chimeras are mated with female and MF1 and 129 mice, and germline transmission is determined by the agouti coat color and by PCR genotyping respectively.

10 ANTIBODIES

- The present invention further provides for antibodies which bind to a Tarzan polypeptide, fragment, homologue, variant or derivative thereof. Particularly Tarzan expression, and in particular in diagnosing a Tarzan GPCR associated disease. Other preferred antibodies include those which have therapeutic activity, i.e., which may be used in a therapeutic manner to treat, manage or prevent any Tarzan GPCR associated disease.

- For the purposes of this invention, the term "antibody", unless specified to the contrary, includes but is not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library. Such fragments include
20 fragments of whole antibodies which retain their binding activity for a target substance, Fv, F(ab') and F(ab')₂ fragments, as well as single chain antibodies (scFv), fusion proteins and other synthetic proteins which comprise the antigen-binding site of the antibody. The antibodies and fragments thereof may be humanised antibodies, for example as described in EP-A-239400. Furthermore, antibodies with fully human variable regions (or their
25 fragments), for example, as described in US Patent Nos. 5,545,807 and 6,075,181 may also be used. Neutralizing antibodies, i.e., those which inhibit biological activity of the substance amino acid sequences, are especially preferred for diagnostics and therapeutics.

Antibodies may be produced by standard techniques, such as by immunisation or by using a phage display library.

A polypeptide or peptide of the present invention may be used to develop an antibody by known techniques. Such an antibody may be capable of binding specifically
5 to the Tarzan GPCR protein or homologue, fragment, etc.

If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) may be immunised with an immunogenic composition comprising a polypeptide or peptide of the present invention. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but
10 are not limited to, Freund's, mineral gels such as aluminium hydroxide, and surface active substances such as lyssolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (*Bacilli Calmette-Guerin*) and *Corynebacterium parvum* are potentially useful human adjuvants which may be employed if purified the substance amino acid sequence is administered to immunologically
15 compromised individuals for the purpose of stimulating systemic defence.

Serum from the immunised animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to an epitope obtainable from a polypeptide of the present invention contains antibodies to other antigens, the polyclonal antibodies can be purified by immunoaffinity chromatography. Techniques for producing
20 and processing polyclonal antisera are known in the art. In order that such antibodies may be made, the invention also provides amino acid sequences of the invention or fragments thereof haptenised to another amino acid sequence for use as immunogens in animals or humans.

Monoclonal antibodies directed against epitopes obtainable from a polypeptide or
25 peptide of the present invention can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or

transfection with Epstein-Barr virus. Panels of monoclonal antibodies produced against orbit epitopes can be screened for various properties; i.e., for isotype and epitope affinity.

Monoclonal antibodies may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, 5 but are not limited to, the hybridoma technique originally described by Kohler and Milstein (1975 *Nature* 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kosbor *et al* (1983) *Immunol Today* 4:72; Cote *et al* (1983) *Proc Natl Acad Sci* 80:2026-2030) and the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, pp. 77-96, Alan R. Liss, Inc., 1985).

10 In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison *et al* (1984) *Proc Natl Acad Sci* 81:6851-6855; Neuberger *et al* (1984) *Nature* 312:604-608; Takeda *et al* (1985) *Nature* 314:452-454). Alternatively, techniques described for the production of 15 single chain antibodies (US Patent No. 4,946,779) can be adapted to produce the substance specific single chain antibodies.

Antibodies, both monoclonal and polyclonal, which are directed against epitopes obtainable from a polypeptide or peptide of the present invention are particularly useful in diagnosis, and those which are neutralising are useful in passive immunotherapy.

20 Monoclonal antibodies, in particular, may be used to raise anti-idiotype antibodies. Anti-idiotype antibodies are immunoglobulins which carry an "internal image" of the substance and/or agent against which protection is desired. Techniques for raising anti-idiotype antibodies are known in the art. These anti-idiotype antibodies may also be useful in therapy.

25 Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in Orlandi *et al* (1989, *Proc Natl Acad Sci* 86: 3833-3837), and Winter G and Milstein C (1991; *Nature* 349:293-299).

Antibody fragments which contain specific binding sites for the polypeptide or peptide may also be generated. For example, such fragments include, but are not limited to, the F(ab')₂ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse WD *et al* (1989) Science 256:1275-1281).

- Techniques for the production of single chain antibodies (U.S. Pat. No. 4,946,778) can also be adapted to produce single chain antibodies to polypeptides of this invention.
- 10 Also, transgenic mice, or other organisms including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

- 15 Antibodies against Tarzan GPCR polypeptides may also be employed to treat infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; diabetes, obesity; anorexia; bulimia; asthma; parkinson's disease; thrombosis; acute heart failure; hypotension; hypertension; erectile dysfunction; urinary retention; metabolic bone diseases such as osteoporosis and osteo petrosis; angina pectoris; myocardial infarction; ulcers; asthma; allergies; rheumatoid
- 20 arthritis; inflammatory bowel disease; irritable bowel syndrome benign prostatic hypertrophy; and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles dela Tourett's syndrome

DIAGNOSTIC ASSAYS

- 25 This invention also relates to the use of Tarzan GPCR polynucleotides and polypeptides (as well as homologues, variants and derivatives thereof) for use in diagnosis as diagnostic reagents or in genetic analysis. Nucleic acids complementary to or capable of

hybridising to Tarzan GPCR nucleic acids (including homologues, variants and derivatives), as well as antibodies against Tarzan polypeptides are also useful in such assays.

We provide for a natural variant of Tarzan polypeptide or nucleic acid, and the use
5 of such a natural variant in diagnosis of Tarzan associated disease. Tarzan polymorphisms may include differences at the nucleic acid level, which may or may not reflect differences in the amino acid level. Preferably, such Tarzan variants or mutants are such that they include changes in the amino acid level. However, the invention also encompasses Tarzan polymorphisms which occur in non-coding regions, for example, expression control
10 regions such as promoters and enhancers.

Polymorphisms in Tarzan include deletions of one or more nucleic acids, insertions of one or more nucleic acids, inversions, etc. Preferably, Tarzan polymorphisms comprise single nucleotide polymorphisms.

Polymorphisms in Tarzan may be identified by comparing sequences at the
15 appropriate level (whether nucleic acid or protein) between individuals in a population. Differences in sequences may be reflected in different physical properties, and techniques for detecting these may rely on detection of changes in physical properties. For example, single nucleotide polymorphisms may be detected as restriction fragment length polymorphisms (i.e., difference in susceptibility to digestion by a restriction enzyme).
20 Furthermore, SNPs may affect the migration or mobility of a nucleic acid fragment or protein fragment in a gel.

Non-coding polymorphisms in Tarzan may be identified by sequencing non-coding regions of Tarzan. For example, control regions of the Tarzan gene, such as enhancers and promoters may be sequenced to identify polymorphisms. The effect of such non-coding
25 polymorphisms on the expression level of Tarzan may be determined by constructing transgenic mice (as described below) comprising the mutant Tarzan sequences, or by generating expression constructs and transfection into cell lines. In each case, the expression level of Tarzan is detected, by RT-PCR or antibody Western staining, to

determine the effect of the mutation in the control of expression of Tarzan. Useful Tarzan polymorphisms are those which modulate the level of expression, whether by up-regulation or down-regulation of Tarzan levels.

Accordingly, this invention provides for a variant or mutant or polymorphism in a
5 non-coding region of Tarzan, preferably in a control region of Tarzan, preferably in a promoter and/or enhancer of Tarzan, which is capable of modulating the level of expression of Tarzan in an organism. The invention also provides for a set of two or more of such mutants or variants or polymorphisms, preferably non-coding polymorphisms. The invention also provides for the use of such variants or polymorphisms or sets of variants to
10 identify nucleic acid and/or amino acid positions, in which changes to such positions affect the level of expression of Tarzan. The invention also provides for a transgenic animal comprising a variant or mutant or polymorphism of Tarzan, preferably, a non-coding polymorphism.

Detection of a mutated form of the Tarzan GPCR gene associated with a
15 dysfunction will provide a diagnostic tool that can add to or define a diagnosis of a disease or susceptibility to a disease which results from under-expression, over-expression or altered expression of Tarzan GPCR. Individuals carrying mutations in the Tarzan GPCR gene (including control sequences) may be detected at the DNA level by a variety of techniques.

20 For example, DNA may be isolated from a patient and the DNA polymorphism pattern of Tarzan determined. The identified pattern is compared to controls of patients known to be suffering from a disease associated with over-, under- or abnormal expression of Tarzan. Patients expressing a genetic polymorphism pattern associated with Tarzan associated disease may then be identified. Genetic analysis of the Tarzan GPCR gene may
25 be conducted by any technique known in the art. For example, individuals may be screened by determining DNA sequence of a Tarzan allele, by RFLP or SNP analysis, etc. Patients may be identified as having a genetic predisposition for a disease associated with the over-, under-, or abnormal expression of Tarzan by detecting the presence of a DNA polymorphism in the gene sequence for Tarzan or any sequence controlling its expression.

Patients so identified can then be treated to prevent the occurrence of Tarzan associated disease, or more aggressively in the early stages of Tarzan associated disease to prevent the further occurrence or development of the disease. Tarzan associated diseases include any one of trigeminal neuralgia, orofacial pain, pain associated with toothache, 5 irritable bowel syndrome, Barrett's oesophagus, glaucoma, pain associated with cancer, diabetic neuropathies, Herpes infections, HIV infections, migraine and skin sensitivity associated with migraine, allodynia, toothache, neuroma (whether caused by amputation, nerve transaction or trauma), nerve compression (caused by tumours, entrapment or crush), pain due to damage of the spinal cord or brain; dementia, dyslexia, dyskinesias, 10 tremor, Parkinson's, benign essential tremor, chorea, epilepsy and ballismus, for example occurring through stroke, trauma, degeneration or malignancy; cystic fibrosis and hyperactive bladder.

The present invention further discloses a kit for the identification of a patient's genetic polymorphism pattern associated with Tarzan associated disease. The kit includes 15 DNA sample collecting means and means for determining a genetic polymorphism pattern, which is then compared to control samples to determine a patient's susceptibility to Tarzan associated disease. Kits for diagnosis of a Tarzan associated disease comprising Tarzan polypeptide and/or an antibody against such a polypeptide (or fragment of it) are also provided.

20 Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. In a preferred embodiment, the DNA is obtained from blood cells obtained from a finger prick of the patient with the blood collected on absorbent paper. In a further preferred embodiment, the blood is collected on an AmpliCard.TM. (University of Sheffield, Department of Medicine and 25 Pharmacology, Royal Hallamshire Hospital, Sheffield, England S10 2JF).

The DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. Oligonucleotide DNA primers that target the specific polymorphic DNA region within the genes of interest may be prepared so that in the PCR reaction amplification of the target sequences is achieved.

RNA or cDNA may also be used as templates in similar fashion. The amplified DNA sequences from the template DNA may then be analyzed using restriction enzymes to determine the genetic polymorphisms present in the amplified sequences and thereby provide a genetic polymorphism profile of the patient. Restriction fragments lengths may

5 be identified by gel analysis. Alternatively, or in conjunction, techniques such as SNP (single nucleotide polymorphisms) analysis may be employed.

Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled Tarzan GPCR nucleotide sequences. Perfectly

10 matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, e.g., Myers et al, *Science* (1985)230:1242. Sequence changes at specific locations may also be revealed by nuclease

15 protection assays, such as RNase and S1protection or the chemical cleavage method. See Cotton et al., *Proc Natl Acad Sci USA* (1985) 85: 4397-4401. In another embodiment, an array of oligonucleotides probes comprising the Tarzan GPCR nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and

20 can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability. (See for example: M.Chee et al., *Science*, Vol 274, pp 610-613 (1996)).

Single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc Natl. Acad. Sci USA:* 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; and Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control Tarzan nucleic acids may be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even

25 a single base change. The DNA fragments may be labeled or detected with labeled probes.

The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al.

5 (1991) Trends Genet 7:5).

The diagnostic assays offer a process for diagnosing or determining a susceptibility to Tarzan associated diseases, for example, infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; diabetes, obesity; anorexia; bulimia; asthma; parkinson's disease; thrombosis; acute heart failure; hypotension; hypertension; erectile dysfunction; urinary retention; metabolic bone diseases such as osteoporosis and osteo petrosis; angina pectoris; myocardial infarction; 10 ulcers; asthma; allergies; rheumatoid arthritis; inflammatory bowel disease; irritable bowel syndrome benign prostatic hypertrophy; and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles dela Tourett's 15 syndrome.

The presence of Tarzan GPCR polypeptides and nucleic acids may be detected in a sample. Thus, infections and diseases as listed above can be diagnosed by methods comprising determining from a sample derived from a subject an abnormally decreased or 20 increased level of the Tarzan GPCR polypeptide or Tarzan GPCR mRNA. The sample may comprise a cell or tissue sample from an organism suffering or suspected to be suffering from a disease associated with increased, reduced or otherwise abnormal Tarzan GPCR expression, including spatial or temporal changes in level or pattern of expression. The level or pattern of expression of Tarzan in an organism suffering from or suspected to 25 be suffering from such a disease may be usefully compared with the level or pattern of expression in a normal organism as a means of diagnosis of disease.

In general therefore, the invention includes a method of detecting the presence of a nucleic acid comprising a Tarzan GPCR nucleic acid in a sample, by contacting the sample with at least one nucleic acid probe which is specific for said nucleic acid and

- monitoring said sample for the presence of the nucleic acid. For example, the nucleic acid probe may specifically bind to the Tarzan GPCR nucleic acid, or a portion of it, and binding between the two detected; the presence of the complex itself may also be detected. Furthermore, the invention encompasses a method of detecting the presence of a Tarzan GPCR polypeptide by contacting a cell sample with an antibody capable of binding the polypeptide and monitoring said sample for the presence of the polypeptide. This may conveniently be achieved by monitoring the presence of a complex formed between the antibody and the polypeptide, or monitoring the binding between the polypeptide and the antibody. Methods of detecting binding between two entities are known in the art, and include FRET (fluorescence resonance energy transfer), surface plasmon resonance, etc.

Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as a Tarzan GPCR, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

The present invention relates to a diagnostic kit for a disease or susceptibility to a Tarzan associated disease (including an infection), for example, infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; diabetes; obesity; anorexia; bulimia; asthma; parkinson's disease; thrombosis; acute heart failure; hypotension; hypertension; erectile dysfunction; urinary retention; metabolic bone diseases such as osteoporosis and osteo petrosis; angina pectoris; myocardial infarction; ulcers; asthma; allergies; rheumatoid arthritis; inflammatory bowel disease; irritable bowel syndrome benign prostatic hypertrophy; and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles dela Tourett's syndrome.

The diagnostic kit comprises a Tarzan GPCR polynucleotide or a fragment thereof; a complementary nucleotide sequence; a Tarzan GPCR polypeptide or a fragment thereof, or an antibody to a Tarzan GPCR polypeptide.

CHROMOSOME ASSAYS

5 The nucleotide sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. As described above, human Tarzan GPCR is found to map to homo sapiens chromosome 13q32.2.

10 The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, Mendelian heritance in Man (available on line through Johns Hopkins University Welch Medical Library). The 15 relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

20 The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

PROPHYLACTIC AND THERAPEUTIC METHODS

This invention provides methods of treating an abnormal conditions related to both an excess of and insufficient amounts of Tarzan GPCR activity.

If the activity of Tarzan GPCR is in excess, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as hereinabove described along with a pharmaceutically acceptable carrier in an amount effective to inhibit activation by blocking binding of ligands to the Tarzan GPCR, or by 5 inhibiting a second signal, and thereby alleviating the abnormal condition.

In another approach, soluble forms of Tarzan GPCR polypeptides still capable of binding the ligand in competition with endogenous Tarzan GPCR may be administered. Typical embodiments of such competitors comprise fragments of the Tarzan GPCR polypeptide.

10 In still another approach, expression of the gene encoding endogenous Tarzan GPCR can be inhibited using expression blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or separately administered. See, for example, O'Connor, *J Neurochem* (1991) 56:560 in *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, Fla. (1988). Alternatively, oligonucleotides which form triple helices with the gene 15 can be supplied. See, for example, Lee et al., *Nucleic Acids Res* (1979) 6:3073; Cooney et al., *Science* (1988) 241:456; Dervan et al., *Science* (1991) 251:1360. These oligomers can be administered per se or the relevant oligomers can be expressed *in vivo*.

For treating abnormal conditions related to an under-expression of Tarzan GPCR 20 and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates Tarzan GPCR, i.e., an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of 25 Tarzan GPCR by the relevant cells in the subject. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces

infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo*. For overview of gene therapy, see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human

- 5 Molecular Genetics, T Strachan and A P Read, BIOS Scientific Publishers Ltd (1996).

FORMULATION AND ADMINISTRATION

Peptides, such as the soluble form of Tarzan GPCR polypeptides, and agonists and antagonist peptides or small molecules, may be formulated in combination with a suitable pharmaceutical carrier. Such formulations comprise a therapeutically effective amount of

- 10 the polypeptide or compound, and a pharmaceutically acceptable carrier or excipient. Such carriers include but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. Formulation should suit the mode of administration, and is well within the skill of the art. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of
15 the aforementioned compositions of the invention.

Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

Preferred forms of systemic administration of the pharmaceutical compositions include injection, typically by intravenous injection. Other injection routes, such as

- 20 subcutaneous, intramuscular, or intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if properly formulated in enteric or encapsulated formulations, oral administration may also be possible. Administration of these compounds may also be topical and/or localize, in the
25 form of salves, pastes, gels and the like.

The dosage range required depends on the choice of peptide, the route of administration, the nature of the formulation, the nature of the subject's condition, and the

judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 µg/kg of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require 5 higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for 10 example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide ex vivo, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

PHARMACEUTICAL COMPOSITIONS

The present invention also provides a pharmaceutical composition comprising 15 administering a therapeutically effective amount of the polypeptide, polynucleotide, peptide, vector or antibody of the present invention and optionally a pharmaceutically acceptable carrier, diluent or excipients (including combinations thereof).

The pharmaceutical compositions may be for human or animal usage in human and veterinary medicine and will typically comprise any one or more of a pharmaceutically 20 acceptable diluent, carrier, or excipient. Acceptable carriers or diluents for therapeutic use are well known in the pharmaceutical art, and are described, for example, in Remington's Pharmaceutical Sciences, Mack Publishing Co. (A. R. Gennaro edit. 1985). The choice of pharmaceutical carrier, excipient or diluent can be selected with regard to the intended route of administration and standard pharmaceutical practice. The pharmaceutical 25 compositions may comprise as - or in addition to - the carrier, excipient or diluent any suitable binder(s), lubricant(s), suspending agent(s), coating agent(s), solubilising agent(s).

Preservatives, stabilizers, dyes and even flavoring agents may be provided in the pharmaceutical composition. Examples of preservatives include sodium benzoate, sorbic acid and esters of p-hydroxybenzoic acid. Antioxidants and suspending agents may be also used.

- 5 There may be different composition/formulation requirements dependent on the different delivery systems. By way of example, the pharmaceutical composition of the present invention may be formulated to be delivered using a mini-pump or by a mucosal route, for example, as a nasal spray or aerosol for inhalation or ingestable solution, or parenterally in which the composition is formulated by an injectable form, for delivery, 10 by, for example, an intravenous, intramuscular or subcutaneous route. Alternatively, the formulation may be designed to be delivered by both routes.

Where the agent is to be delivered mucosally through the gastrointestinal mucosa, it should be able to remain stable during transit though the gastrointestinal tract; for example, it should be resistant to proteolytic degradation, stable at acid pH and resistant to 15 the detergent effects of bile.

Where appropriate, the pharmaceutical compositions can be administered by inhalation, in the form of a suppository or pessary, topically in the form of a lotion, solution, cream, ointment or dusting powder, by use of a skin patch, orally in the form of tablets containing excipients such as starch or lactose, or in capsules or ovules either alone 20 or in admixture with excipients, or in the form of elixirs, solutions or suspensions containing flavouring or colouring agents, or they can be injected parenterally, for example intravenously, intramuscularly or subcutaneously. For parenteral administration, the compositions may be best used in the form of a sterile aqueous solution which may contain other substances, for example enough salts or monosaccharides to make the 25 solution isotonic with blood. For buccal or sublingual administration the compositions may be administered in the form of tablets or lozenges which can be formulated in a conventional manner.

VACCINES

- Another embodiment of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with the Tarzan GPCR polypeptide, or a fragment thereof, adequate to produce antibody and/or T cell immune response to protect said animal from infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; diabetes; obesity; anorexia; bulimia; asthma; parkinson's disease; thrombosis; acute heart failure; hypotension; hypertension; erectile dysfunction; urinary retention; metabolic bone diseases such as osteoporosis and osteo petrosis; angina pectoris; myocardial infarction; ulcers; asthma; allergies; rheumatoid arthritis; inflammatory bowel disease; irritable bowel syndrome benign prostatic hypertrophy; and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles dela Tourett's syndrome, among others.
- Yet another embodiment of the invention relates to a method of inducing immunological response in a mammal which comprises delivering a Tarzan GPCR polypeptide via a vector directing expression of a Tarzan GPCR polynucleotide in vivo in order to induce such an immunological response to produce antibody to protect said animal from diseases.
- A further embodiment of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces an immunological response in that mammal to a Tarzan GPCR polypeptide wherein the composition comprises a Tarzan GPCR polypeptide or Tarzan GPCR gene. The vaccine formulation may further comprise a suitable carrier.
- Since the Tarzan GPCR polypeptide may be broken down in the stomach, it is preferably administered parenterally (including subcutaneous, intramuscular, intravenous, intradermal etc. injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants,

buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a

- 5 freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

- 10 Vaccines may be prepared from one or more polypeptides or peptides of the present invention.

The preparation of vaccines which contain an immunogenic polypeptide(s) or peptide(s) as active ingredient(s), is known to one skilled in the art. Typically, such vaccines are prepared as injectables, either as liquid solutions or suspensions; solid forms

- 15 suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations
20 thereof.

In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted

from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion.

Further examples of adjuvants and other agents include aluminum hydroxide, aluminum phosphate, aluminum potassium sulfate (alum), beryllium sulfate, silica, kaolin, 5 carbon, water-in-oil emulsions, oil-in-water emulsions, muramyl dipeptide, bacterial endotoxin, lipid X, *Corynebacterium parvum* (*Propionobacterium acnes*), *Bordetella pertussis*, polyribonucleotides, sodium alginate, lanolin, lysolecithin, vitamin A, saponin, liposomes, levamisole, DEAE-dextran, blocked copolymers or other synthetic adjuvants. Such adjuvants are available commercially from various sources, for example, Merck 10 Adjuvant 65 (Merck and Company, Inc., Rahway, N.J.) or Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, Michigan).

Typically, adjuvants such as Amphigen (oil-in-water), Alhydrogel (aluminum hydroxide), or a mixture of Amphigen and Alhydrogel are used. Only aluminum hydroxide is approved for human use.

15 The proportion of immunogen and adjuvant can be varied over a broad range so long as both are present in effective amounts. For example, aluminum hydroxide can be present in an amount of about 0.5% of the vaccine mixture (Al₂O₃ basis). Conveniently, the vaccines are formulated to contain a final concentration of immunogen in the range of from 0.2 to 200 µg/ml, preferably 5 to 50 µg/ml, most preferably 15 µg/ml.

20 After formulation, the vaccine may be incorporated into a sterile container which is then sealed and stored at a low temperature, for example 4°C, or it may be freeze-dried. Lyophilisation permits long-term storage in a stabilised form.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are 25 suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures

containing the active ingredient in the range of 0.5% to 10%, preferably 1% to 2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions,
5 suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10% to 95% of active ingredient, preferably 25% to 70%. Where the vaccine composition is lyophilised, the lyophilised material may be reconstituted prior to administration, e.g. as a suspension. Reconstitution is preferably effected in buffer

10 Capsules, tablets and pills for oral administration to a patient may be provided with an enteric coating comprising, for example, Eudragit "S", Eudragit "L", cellulose acetate, cellulose acetate phthalate or hydroxypropylmethyl cellulose.

15 The polypeptides of the invention may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric and maleic. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine and procaine.

20 ADMINISTRATION

Typically, a physician will determine the actual dosage which will be most suitable for an individual subject and it will vary with the age, weight and response of the particular patient. The dosages below are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited.

25 The pharmaceutical and vaccine compositions of the present invention may be administered by direct injection. The composition may be formulated for parenteral, mucosal, intramuscular, intravenous, subcutaneous, intraocular or transdermal

administration. Typically, each protein may be administered at a dose of from 0.01 to 30 mg/kg body weight, preferably from 0.1 to 10 mg/kg, more preferably from 0.1 to 1 mg/kg body weight.

The term "administered" includes delivery by viral or non-viral techniques. Viral
5 delivery mechanisms include but are not limited to adenoviral vectors, adeno-associated viral (AAV) vectors, herpes viral vectors, retroviral vectors, lentiviral vectors, and baculoviral vectors. Non-viral delivery mechanisms include lipid mediated transfection, liposomes, immunoliposomes, lipofectin, cationic facial amphiphiles (CFAs) and combinations thereof. The routes for such delivery mechanisms include but are not limited to mucosal, nasal,
10 oral, parenteral, gastrointestinal, topical, or sublingual routes.

The term "administered" includes but is not limited to delivery by a mucosal route, for example, as a nasal spray or aerosol for inhalation or as an ingestable solution; a parenteral route where delivery is by an injectable form, such as, for example, an intravenous, intramuscular or subcutaneous route.

15 The term "co-administered" means that the site and time of administration of each of for example, the polypeptide of the present invention and an additional entity such as adjuvant are such that the necessary modulation of the immune system is achieved. Thus, whilst the polypeptide and the adjuvant may be administered at the same moment in time and at the same site, there may be advantages in administering the polypeptide at a
20 different time and to a different site from the adjuvant. The polypeptide and adjuvant may even be delivered in the same delivery vehicle - and the polypeptide and the antigen may be coupled and/or uncoupled and/or genetically coupled and/or uncoupled.

The polypeptide, polynucleotide, peptide, nucleotide, antibody of the invention and optionally an adjuvant may be administered separately or co-administered to the host
25 subject as a single dose or in multiple doses.

The vaccine composition and pharmaceutical compositions of the present invention may be administered by a number of different routes such as injection (which includes

parenteral, subcutaneous and intramuscular injection) intranasal, mucosal, oral, intra-vaginal, urethral or ocular administration.

The vaccines and pharmaceutical compositions of the present invention may be conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For 5 suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, may be 1% to 2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of 10 mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10% to 95% of active ingredient, preferably 25% to 70%. Where the vaccine composition is 15 lyophilised, the lyophilised material may be reconstituted prior to administration, e.g. as a suspension. Reconstitution is preferably effected in buffer.

EXAMPLES

Example 1. Transgenic TARZAN Knock-Out Mouse

Construction of Tarzan Gene Targeting Vector

Murine Tarzan genomic clones are isolated from a mouse large insert PAC library obtained from HGMP (Hinxton, UK) using a probe sequence amplified from the predicted murine open reading frame cDNA sequence (SEQ ID NO: 4), using standard techniques. The primers musTarzan F1 and R1 are used to generate the probe. 20

The isolated murine Tarzan genomic clones are then restriction mapped in the region of the Tarzan gene using small oligonucleotide probes and standard techniques. The murine genomic locus is partially sequenced to enable the design of homologous arms to 25

clone into the targeting vector. The structure of the targeting vector, including its restriction sites, is shown in Figure 5.

The murine Tarzan gene has a single coding exon. The targeting strategy is designed to remove the majority of this exon. A 1.1 kb 5' homologous arm and a 3.7 kb 3' homologous arm flanking the coding exon are amplified by PCR and the fragments are cloned into the targeting vector. The 5' end of each oligonucleotide primer used to amplify the arms is synthesized to contain a different recognition site for a rare-cutting restriction enzyme, compatible with the cloning sites of the vector polylinkers and absent from the arms themselves. In the case of Tarzan, the primers are designed as listed in the sequence table below, with 5' arm cloning enzymes of AgeI/NotI and 3' arm cloning enzymes of Ascl/FseI.

In addition to the arm primer pairs (5'armF/5'armR and 3'armF/3'armR), further primers specific to the Tarzan locus are designed for the following purposes: 5' and 3' probe primer pairs (5'prF/5'prR and 3'prF/3'prR) to amplify two short 200-300bp fragments of non-repetitive genomic DNA external to and extending beyond each arm, to allow Southern analysis of the targeted locus, in isolated putative targeted clones; a mouse genotyping primer pair (hetF and hetR) which allows differentiation between wild-type, heterozygote and homozygous mice, when used in a multiplex PCR with a vector specific primer, in this case, Asc306; and lastly, a target screening primer (5'scr.2) which anneals upstream of the end of the 5' arm region, and which produces a target event specific 1.2kb amplimer when paired with a primer specific to the 5' end of the vector (DR1). This amplimer can only be derived from template DNA from cells where the desired genomic alteration has occurred and allows the identification of correctly targeted cells from the background of clones containing randomly integrated copies of the vector. The location of these primers and the genomic structure of the TARZAN locus used in the targeting strategy is shown in SEQ ID NO: 10.

musTarzan 5'prF	TTTGATGGAGTAGATTCACTGTGTCAGAC
musTarzan 5'prR	GTCGGAGAGCTTGGGAATCCAAAC
musTarzan 5'armF.2 Age	tttaccggTTCCCTACATTAGCGACTTTCTGGAC
musTarzan 5'armR Not	aaagcgccgCCAGGAAATCCGAATCACTGGCTGGAC
musTarzan 3'armF Asc	aaaggcgccgcGATGCAAAGCCAGTGGGGACCTTGAAC
musTarzan 3'armR Fse	tttggccGGcCTGAGCATGGCAGTTCCCTTG

musTarzan 5'scr.2	AATCATCCTTCTTGCTGGGTTAGAC
musTarzan 3'prF	CGTGGCACTCAAGTCTGGATGCACCTG
musTarzan 3'prR	AGACACTAGGACATTAATCCCTGTGAC
musTarzan hetF	TCCACATCTTGAGGGTCATTGGATCG
musTarzan hetR	TTCTTCCCTGTTCAAGGTCCCCACTG
musTarzan F1	GGATTTCCCTGGATTACCCAAGTGC
musTarzan R1	CAGATACAGCAAGTCCGCAAGGC
Asc306	AATGGCCGCTTTCTGGATTATCGAC
DR1	CATGCCGCCTGCGCCCTATTGATCATG

Table 1. Tarzan Primer Sequences

The position of the homology arms is chosen to functionally disrupt the Tarzan gene by deleting the seven transmembrane spanning regions. A targeting vector is prepared where the deleted Tarzan sequence is replaced with non-homologous sequences 5 composed of an endogenous gene expression reporter (a frame independent lacZ gene) upstream of a selection cassette composed of a promoted neomycin phosphotransferase (neo) gene arranged in the same orientation as the Tarzan gene.

Once the 5' and 3' homology arms had been cloned into the targeting vector pTKSIBLMNL (see Figure 5), a large highly pure DNA preparation is made using 10 standard molecular biology techniques. 20 µg of the freshly prepared endotoxin free DNA is restricted with another rare-cutting restriction enzyme PmeI, present at a unique site in the vector backbone between the ampicillin resistance gene and the bacterial origin of replication. The linearized DNA is then precipitated and resuspended in 100 µl of Phosphate Buffered Saline, ready for electroporation.

15 24 hours following electroporation the transfected cells are cultured for 9 days in medium containing 200 µg/ml neomycin. Clones are picked into 96 well plates, replicated and expanded before being screened by PCR (using primers 5'scr.2 and DR1, as described above) to identify clones in which homologous recombination had occurred between the endogenous Tarzan gene and the targeting construct. Positive clones can be identified at a 20 rate of 1 to 5%. These clones are expanded to allow replicas to be frozen and sufficient high quality DNA to be prepared for Southern blot confirmation of the targeting event

using the external 5' and 3' probes prepared as described above, all using standard procedures (Russ et al, 2000, *Nature* 2000 Mar 2; 404(6773):95-9).

The structure of the genomic locus of mouse Tarzan before knock-out is depicted in Figure 3. The structure of the genomic locus of mouse Tarzan after knock-out is
5 depicted in Figure 4.

Generation of Tarzan GPCR Deficient Mice

C57BL/6 female and male mice are mated and blastocysts are isolated at 3.5 days of gestation. 10-12 cells from a chosen clone are injected per blastocyst and 7-8 blastocysts are implanted in the uterus of a pseudopregnant F1 female. A litter of chimeric
10 pups are born containing several high level (up to 100%) agouti males (the agouti coat colour indicates the contribution of cells descended from the targeted clone). These male chimeras are mated with female and MF1 and 129 mice, and germline transmission is determined by the agouti coat colour and by PCR genotyping respectively.

Genotyping is carried out by PCR on lysed tail clips, using the primers hetF and
15 hetR with a third, vector specific primer (Asc306). This multiplex PCR allows amplification from the wild-type locus (if present) from primers hetF and hetR giving a 220 bp band. The site for hetF is deleted in the knockout mice, so this amplification will fail from a targeted allele. However, the Asc306 primer will amplify a 343 bp band from the targeted locus, in combination with the hetR primer which anneals to a region just
20 inside the 3' arm. Therefore, this multiplex PCR reveals the genotype of the litters as follows: wild-type samples will exhibit a single 220 bp band; heterozygous DNA samples yield two bands at 220 bp and 343 bp; and the homozygous samples will show only the target specific 343 bp band.

LacZ Staining

25 The X gal staining of dissected tissues is performed in the following manner.

- Representative tissue slices are made of large organs. Whole small organs and tubes are sliced open, so fixative and stain will penetrate. Tissues are rinsed thoroughly in PBS (phosphate buffered saline) to remove blood or gut contents. Tissues are placed in fixative (PBS containing 2% formaldehyde, 0.2% glutaraldehyde, 0.02% NP40, 1mM MgCl₂, Sodium deoxycholate 0.23mM) for 30-45 minutes. Following three 5 minute washes in PBS, tissues are placed in Xgal staining solution (4mM K Ferrocyanide, 4mMKFerricyanide, 2mM MgCl₂, 1mg/mlX-gal in PBS) for 18 hours at 30C. Tissues are PBS washed 3 times, postfixed for 24 hours in 4% formaldehyde, PBS washed again before storage in 70% ethanol.
- 5 10 To identify Xgal stained tissues, dehydrated tissues are wax embedded, and 7um section sections cut, counterstained with 0.01% Safranin (9-10 min).

Behavioural and Neurological Testing

Mice are housed under a 12h lights-12h dark light schedule (lights-on at 6 am) with free access to food and water. Mice (n=12), of mixed sexes, aged 3 to 4 months old, 15 are submitted to behavioural testing during the morning, between 10h and 13h to avoid any differential circadian effect on the test results.

Example 2. Expression of Recombinant TARZAN Protein

Recombinant TARZAN is expressed and purified. Two systems are used for expression.

20 ***pTOPO-Echo Donor Based Construct***

A polynucleotide having the sequence shown in SEQ ID NO: 2 is obtained from the human TARZAN nucleic acid sequence (SEQ ID NO: 1). The SEQ ID NO: 2 polynucleotide is cloned into a pTOPO-Echo Donor vector module (Invitrogen pUniV5/His Cat# ET001-10). Transfection of the resulting construct into a host strain and 25 induction of expression (according to the manufacturer's instructions) yields a fusion protein having the sequence of SEQ ID NO: 6.

The fusion polypeptide SEQ ID NO: 6 contains a C terminal V5 tag (residues 344 to 357) and His tag (residues 358 to 363) to aid detection and purification.

pCDNA5-JE Based Construct

- A polynucleotide having the sequence shown in SEQ ID NO: 7 is amplified by
- 5 PCR using the oligonucleotide primers
AATAAAGCTAGCACCATGAATGAGCCACTAGAC and
ATAATCTCGAGTCAAGGGTTGTTGAGTAACTAATTTC to incorporate new
restriction sites, NheI and XhoI at the 5-prime and 3-prime ends respectively of Tarzan.
This is then digested and ligated into similarly digested pcDNA5-JE (Invitrogen Cat# -
10 K6010-01 vector modified to remove BGH Poly-A).

The resulting construct is used for high level expression in CHO-K1 cells , and other mammalian cell lines, under the control of the *cmy* promoter to yield a native polypeptide.

- A polynucleotide having the sequence shown in SEQ ID NO: 2 is amplified by
- 15 PCR using the oligonucleotide primers
AATAAAGCTAGCACCATGAATGAGCCACTAGAC and
ATAATCTCGAGTCACTTATCGTCGTACATCCTTGTAAATCAGGGTTGTTGAGTAA
CTAATTTC to incorporate new restriction sites, NheI and XhoI at the 5-prime and 3-prime ends respectively of Tarzan and to include a 3' fusion FLAG tag. This is then
20 digested and ligated into similarly digested pcDNA5-JE (Invitrogen Cat# - K6010-01 vector modified to remove BGH Poly-A).

- The resulting construct is used for high level expression in CHO-K1 cells, and other mammalian cell lines, under the control of the *cmy* promoter to yield a fusion polypeptide with C terminal FLAG tag (double underline) to aid detection and
25 purification. The resultant expressed fusion polypeptide has a sequence shown in SEQ ID NO: 8.

Introduction of Construct into Cells

The expression vector is introduced to the cells by lipofection (using Fugene-6 from Roche, Cat# 1 814 433) among other similar methods.

Both transient and stable transfection of these cells is achieved. In transient expression the cells are transfected by lipofection using a large amount of vector which 5 results in a short-lived fast expression of the protein. In a stable transfection, the vector, which includes a selectable marker for neomycin resistance becomes stably integrated into the genome of the host cell resulting in a long lived cell line with a high expression level of Tarzan.

Cells expressing recombinant TARZAN are used for assay development, antibody 10 production, and other purposes as described.

Expression in Other Host Cells

The recombinant/fusion clone SEQ ID NO: 2 is recombined into a pBAD-Thio-E factor (Invitrogen Cat# ET100-01) for high level bacterial expression under control of the araBAD promoter, using a Cre/Lox mediated recombination system.

15 The recombinant/fusion clone SEQ ID NO: 2 is recombined into a pBlueBac 4.5E (Invitrogen Cat# ET310-01), using a Cre/Lox mediated recombination system, for subsequent recombination into Baculovirus expression systems. Recombination into MaxBac (Invitrogen Cat# K875-02) for high level expression in SF9 and other insect cell lines.

20 The recombinant/fusion clone SEQ ID NO: 2 is recombined into pcDNA 3.1-E (Invitrogen Cat# ET400-01), using a Cre/Lox mediated recombination system, for high level expression in CHO-K1 (Chinese Hamster Ovary) cells, and other mammalian cell lines, under the control of the *cmy* promoter.

Example 3. Expression of TARZAN in the Thyroid and Brain

LacZ staining in Thyroid gland and the brain of the -/- mice can be detected. The thyroid is an essential organ in controlling energy metabolism and this expression supports a role for this target in obesity related disorders.

5 Example 4. Analysis of Blood Samples from TARZAN Knock-Out Mice

Analysis of blood samples taken from mutant TARZAN animals can show an altered level of circulating metabolites and fatty acids. This coupled with the expression seen in the thyroid, may imply a use for the TARZAN protein in the development of therapeutics to treat hypercholesterolaemia, dislipdaemias, obesity or drugs to affect 10 energy regulation.

Example 5. Motor Control and Balance in TARZAN Knock-Out Mouse

TARZAN mutant mice can be used to test the role of the protein in motor co-ordination and balance by their performance on the Rotarod apparatus.

Motor co-ordination and balance are measured by performance on the rotarod. We 15 use an accelerating Rotarod (Ugo Basile, Linton instruments, Jones BJ, Roberts DJ. *The quantitative measurement of motor inco-ordination in naive mice using an accelerating rotarod. J Pharm Pharmacol.* 1968 Apr;20(4):302-4. We also use the Rotarod to assess motor learning by repeating the task over several days.

Mice are placed on the Rotarod, which accelerates at a constant rate from 4 to 40 20 rpm in 5 in. As the Rotarod reaches higher speed, the mice often grip the Rotarod and hang on for a full rotation (i.e. passive rotation). The time at which the mouse makes one full rotation is recorded. Mice are left on the Rotarod after the first passive rotation and allowed to perform the rest of the test until they drop from the rod. Mice are given 3 trials on three consecutive days.

Example 6. Visual Ability, Anxiety and Mobility in TARZAN Knock-Out Mouse

Tests for visual ability, anxiety and mobility in a TARZAN knock-out mouse are conducted using a visual cliff test.

- The Visual cliff is developed by Fox (Fox M.W *Anim Behav*. 1965 Apr 5 Jul;13(2):232-3) and provides a measure of gross visual ability. It evaluates the ability of the mouse to see the drop-off at the edge of a horizontal surface. Time for the animal to move one placed on the cliff (latency) is a measure of anxiety.

- A Perspex box is built with a horizontal plane connected to a vertical drop of 0.5m. A black and white chequerboard pattern accentuates the vertical drop-out. A sheet of clear 10 Perspex is placed across the top horizontal cliff, extending over the top, thus there is the visual appearance of a cliff, but in fact the Perspex provides a solid horizontal surface.

- Each mouse is given 10 trials on the visual cliff. A trial consists of placing the mouse on the centre 'ridge' and noting the time taken from the animal to move off the 'ridge' (latency) and recording the side on which the mouse stepped. The result is 15 considered positive when the animal chooses to walk on the chequered "safe" side and avoids the cliff and negative result for the other way round. After 5 trials the box is turned through 180 degrees and a further 5 trials given. This is done to eliminate the variable of the position of the observer. The platform is cleaned between each animal.

- Example 7. Tests for Sensitivity to External Stimuli and Pain (Analgesia Testing) in 20 TARZAN Knock-Out Mouse: Paw Pressure Test**

TARZAN mutants can be tested for their sensitivity to touch and pain stimuli using a range of tests. The tests in this and the following four Examples assess both neuropathic and inflammatory pain.

- Skin sensitivity is assessed by applying pressure on the hindpaw with a sharpened 25 dowel rod whilst the animal is resting on a grid. Responses are graded as follow: 0 = no

withdrawal; 1 =slow withdrawal of the paw; 2 = medium withdrawal of the paw; 3 = fast withdrawal of the paw.

Example 8. Tests for Sensitivity to External Stimuli and Pain (Analgesia Testing) in TARZAN Knock-Out Mouse: Tail Flick Test

5 A tail flick analgesia test is performed using a Tail-Flick Analgesia Meter. This equipment provides an easy to use method to determine pain sensitivity accurately and reproducibly in rodents (D'Amour and Smith, 1941 *Journal of Pharmacology and Experimental Therapeutics*. Vol 41 p419-424). The instrument has a shutter-controlled lamp as a heat source. The lamp is located below the animal to provide a less confining 10 environment. Tail flick is detected by the automatic detection circuitry, which leaves the user's hands free to handle the animal. The animal is restrained in a ventilated tube and its tail placed on a sensing groove on top of the equipment.

15 Activation of an intense light beam to the tail through opening of the shutter results in discomfort at some point when the animal will flick its tail out of the beam. In the automatic mode a photo-detector detects the tail motion causing the clock to stop and the shutter to close. The total time elapsed between the shutter opening and the animal's 20 reaction is recorded.

Responses of mutant transgenic mice are compared with age and sex matched wild-type mice. A single animal may be subjected to different heat settings to produce an 20 increase in tail temperature no greater than 55°C.

Using the tail flick test TARZAN mutants can be tested for their sensitivity to heat induced pain and compared to their wild-type counterparts. Decreased sensitivity in mutants is associated by the withdrawal of their tails after a longer time period of exposure to the heat source.

Example 9. Tests for Sensitivity to External Stimuli and Pain (Analgesia Testing) in TARZAN Knock-Out Mouse: Formalin Test

The formalin test measures the response to a noxious substances injected into a hind paw. A volume of 20 μ l of a 1% formalin solution is injected through a fine gauge 5 needle subcutaneously into the dorsal surface of one hindpaw. Licking and biting the hindpaw is quantitated as cumulative number of seconds engaged in the behaviours. A rating scale is used: 1= the formalin injected paw rests lightly on the floor bearing less weight; 2= the injected paw is elevated; 3= the injected paw is licked, bitten or shaken.

Two phases of responses are seen in the formalin test. Phase 1 begins immediately 10 after injection and lasts about 10 mins, representing the acute burst of activity from pain fibres. Phase two begins about 20 mins after injection and continues for about one hour. This phase appears to represent responses to tissue damage, including inflammatory hyperalgesia.

Using the formalin test TARZAN mutants may be shown to be less sensitive to 15 inflammatory pain and may show a reduced severity of response in the formalin test when compared to wild type animals.

Example 10. Tests for the Role of TARZAN in Energy Metabolism

TARZAN mutant mice may be used to test the function of the TARZAN protein in energy metabolism by the determination of total body oxygen consumption (VO₂) (a 20 parameter of energy expenditure) through indirect calorimetry. Oxygen is required for combustion of fuels. By measuring total body oxygen consumption the total amount of fuel used can be calculated. Energy expenditure is measured by estimating oxidation rates of macronutrients from rates of respiratory exchange of oxygen and carbon dioxide (Vidal-Puig AJ, Grujic C, Zhang CY, Hagen T, Boss O, Ido Y, Szczepanik A, Wade J, 25 Mootha V, Cortright R: Energy Metabolism in Uncoupling Protein 3 Gene Knockout Mice . J Biol Chem 275:16258-16266, 2000).

These measurements provide information not only about the total energy expenditure but also about preferential use of specific nutrients as a source of energy. Energy expenditure is measured in basal conditions and in response to some established modulators (e.g. fasting would produce a decrease in energy expenditure), specific diet manipulation (e.g. High fat diet should increase energy expenditure), or pharmacological treatment (e.g. adrenergic receptor agonists should induce energy expenditure (8)). This information determines whether energy expenditure is altered in TARZAN mutants and provides insights on the mechanisms involved (e.g. oxidation of fatty acids vs carbohydrates etc). These measurements are performed in calorimeters with individual chambers.

Thus oxygen consumption and respiratory exchange ratio are measured (AA) in basal, fasted (24/36h), and other experimental conditions (e.g. diets, pharmacological treatments) using a calorimeter (e.g. Columbus Instruments, Columbus, OH). Standard conditions of the measurements are settling time 100 sec, measuring time 50 sec, with room air as reference. However, these parameters are optimised specifically for every experiment. Air in the respiratory chamber is continuously exchanged. This chamber has two sensors that measure the difference between the oxygen and CO₂ that enters and leaves the chamber. Experiments may require up to 1-hour calibration time, period during which animals can be placed in the chamber. Thus a single animal is deposited in an individual chamber and left undisturbed during the time required for the experiment. The length of the experiment would be variable (2-9 h) depending on its specific nature (e.g. fed or fasted animals will be in the chamber during 2h while Lipopolisacaride injected animals will require 6-8 hours).

Differences in reading between TARZAN mutant and wild type mice indicate a role for the TARZAN protein in metabolism and indicate the use of this target for the development of therapeutics tackling disorders of hypercholesterolaemia, dislipidaemias and obesity.

Each of the applications and patents mentioned in this document, and each document cited or referenced in each of the above applications and patents, including

during the prosecution of each of the applications and patents ("application cited documents") and any manufacturer's instructions or catalogues for any products cited or mentioned in each of the applications and patents and in any of the application cited documents, are hereby incorporated herein by reference. Furthermore, all documents cited 5 in this text, and all documents cited or referenced in documents cited in this text, and any manufacturer's instructions or catalogues for any products cited or mentioned in this text, are hereby incorporated herein by reference.

Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and 10 spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the claims.

SEQUENCES

SEQ ID NO: 1 Human TARZAN cDNA

hucDNA

5 TTAACCTTATATTTATTAGATGCATTAGTAAC TGCTCATAGTCATTTCTGGAAATTCA
 AATTCCTCACAGGGTCTCTTGAGATTAAGAGAGAGAAGTGGCAAATTAGGATGTTAGA
 ATAATTTCAATTAAAAGTAGATCCTGTTTATTACCCATCATTAATGTTCTGTTCTTCTT
 TATCAGCGAGTTACTGCTCATTTGATTCATATTGCCAAACTGAACCTCTGTTCTTGCAAGAT
 10 GAAAGGAGACAACCAGAATGAGCCACTAGACTATTAGCAAATGCTCTGATTCCCCGATTATG
 CAGCTGCTTTGGAAATTGCACTGATGAAAACATCCCACCAAGATGCACTACCTCCCTGTTATT
 ATGGCATTATCTCCTCGTGGGATTCCAGGAATGCAGTAGTGTATCCACTACATTTCAAAA
 TGAGACCTTGGAAAGAGCAGCACCATCATTATGCTGAACCTGGCTGCACAGATCTGCTGTATCTGA
 CCAGCCTCCCCCTCCTGATTCACTACTATGCCAGTGGCAAACGGATCTTGGAGATTTCATGT
 15 GTAAGTTATCCGCTCAGCTTCATTCAACCTGATAGCAGCATCCTCTCCTCACCTGTTCA
 GCATCTCCGCTACTGTGTGATCATTCAACCAATGAGCTGCTTTCCATTCAACAAACTCGATGTG
 CAGTTGTAGCCTGTGCTGTGGATCATTCACTGGTAGCTGTCAATTCCGATGACCTTCTGTA
 TCACATCAACCAACAGGACCAACAGATCAGCTGTGACCTCACAGTCGGATGAACCTCAATA
 CTATTAAGGGTACAACCTGATTGACTGCAACTACTTCTGCCTCCCTGGTGTAGTGTGACAC
 20 TTGCTATACCACGATTATCCACACTCTGACCCATGGACTGCAAACACTGACAGCTGCCTTAAGCAGA
 AAGCACGAAGGCTAACCAATTCTGCTACTCCTTGCAATTACGTATGTTTTACCCCTCCATATCT
 TGAGGGTCATTGGATCGAATCTGCGCTGTTCAATCAGTTGTCCATTGAGAATCAGATCCATG
 AAGCTTACATCGTTCTAGACCATTAGCTGCTCTGAACACCTTGGTAACCTGTTACTATATGTGG
 TGGTCAGCGACAACCTTCAGCAGGCTGTGCTCAACAGTGTGAGATGCAAAGTAAGCAGGGAACCTTG
 25 AGCAAGCAAAGAAAATTAGTTACTCAAACACCCCTGAAATATTCAATTACTTAACCAAAACAA
 ATACTGCTGATACTTACCTAGCATCCTAAGATGTTAGGATGTCCTCCCTCAATGGAACCTCTGG
 TAAATACTGTGTATTCAAGTAATCATGTGCAAAGCCAGGGCAGAGCTCTAGTTCTGCAATCC
 CTTATTGAGCTCCTCCACTGGGAGATATAAGAATGGGATGCACTGTATATCAGCAAAGTATTCA
 ACATAGTATTACAAGCTATTGAACTCAGAGGCATCTAGAGAACATCTGTTCCACCAACTTACT
 30 ATATATACACGGAAACCAATTCTACCTGCCCCAGATTGCTCAGTAAATTGTGCCAAGATAG
 GAGAAAACCAATTTCACTCATCTTCAATGCTCTGCACCTGGGCCTATTGTATTGAAC
 CATTAGACAATTCAAACCAACTATTGTATCTTCTTAATATTATTTACATCTCAGAGCTCTA
 CAATTGTTCCCTC
 35 <210> 1
 <211> 1179
 <212> DNA
 <213> Homo sapiens
 40 <400> 1
 ggagacaacc atgaatgagc cactagacta ttttagcaa at gcttctgatt tcccccatta 60
 tgcagctgt tttggaaatt gcactgatga aaacatccca ctcagaatgc actacccccc 120
 tgttatttat ggcattatct tcctcggtt atttccaggc aatgcagtag tgatatccac 180
 ttacattttc aaaatgagac ctggaaagag cagcaccatc attatgctga acctggcctg 240
 45 cacagatctg ctgtatctga ccagcctccc cttccctgatt cactactatg ccagtggcga 300
 aaactggatc tttggagatt tcatgtgtaa gtttattccgc ttcaagttcc atttcaacct 360
 gtatagcagc atcccttcc tcacctgttt cagcatcttc cgctactgtg tgatcattca 420
 ccccatgagc tgctttcca ttccacaaaac tcgatgtgca gtttgtagct gtgctgtgt 480
 gtggatcattc tcaactggtag ctgtcattcc gatgaccttc ttgatcacat caaccaacag 540
 50 gaccaacaga tcagcctgtc tcgacccac cagttccggat gaactcaata ctattaagt 600
 gtacaacactg attttgactg caactacttt ctgcctcccc ttggatgatag tgacacttg 660
 ctataccacg attatccaca ctctgaccca tggactgcaa actgacagct gccttaagca 720

5	gaaagcacga aggctaacca ttctgctact ccttgcattt tacgtatgtt ttttaccctt 780 ccatatcttg agggtcattc ggatcgaatc tcgcctgtt tcaatcagtt gttccattga 840 gaatcagatc catgaagctt acatcgttt tagaccatta gctgctctga acaccttgg 900 taacctgtta ctatatgtgg tggtcagcga caactttcag caggctgtct gctcaacagt 960 gagatgaaaa gtaaggcgaga accttggaca agcaaaagaaa attagttact caaacaaccc 1020 ttgaaaatatt tcatttactt aaccaaaaac aaatacttgtc tgatacttta ccttagcatcc 1080 taagatgttc aggatgtctc cctcaatggaa actcctgtta aatactgtgtt attcaagtaa 1140 tcatgtgcca aagccagggc aqagcttcta gttctttgc 1179
---	---

SEQ ID NO: 2 ORF from SEQ ID NO: 1

15 ATGAATGAGCCACTAGACTATTAGCAAATGCTTCTGATTCCCCGATTATGCAGCTGCTTTGGAA
AATTGCACTGATGAAAACATCCCACTCAAGATGCACTACCTCCCTGTTATTATGGCATTATCTTC
CTCGTGGGATTCCAGGCAATGCAGTAGTGATATCCACTACATTCAAATGAGACCTTGGAAAG
AGCAGCACCATCATTATGCTGAACCTGGCTGCACAGATCTGCTGTATCTGACCAGCCTCCCTTC
CTGATTCACTACTATGCCAGTGGCGAAAACGGATCTTGGAGATTTCATGTGTAAGTTATCCGC
TTCAGCTCCATTCAACCTGTATAGCAGCATCCTCTTCCACCTGTTCACTCTCCGCTAC
TGTGTGATCATTCAACCAATGAGCTGCTTTCCATTCAAAAACGATGTGCAAGTGTAGCCTGT
20 GCTGTGGTGGATCATTCACTGGTAGCTGTCATTCCGATGACCTCTTGATCACATCAACCAAC
AGGACCAACAGATCAGCCTGTCGACCTCACCAGTTCGGATGAACCTAACATAACTATAAGTGGTAC
AACCTGATTTCGACTGCAACTACTTCTGCCCTGGTGAAGTGACACTTTGCTATAACUACG
ATATCCACACTCTGACCCATGGACTGCAAACTGACAGCTGCCCTAACAGAAAGCACGAAGGCTA
ACCATTCTGCTACTCCTTGCATTTACGTATGTTTTACCCCTCCATATCTTGAGGGTCATTGG
25 ATCGAATCTGCCCTGTTCAATCAGTTGGCATTGAGAATCAGATCCATGAAGCTTACATCGTT
TCTAGACCATTAGCTGCTCTGAAACACCTTGGTAACCTGTTACTATATGTGGTGGTCAGCGACAAC
TTTCAGCAGGCTGCTGCTCAACAGTGAGATGCAAAGTAAGCGGGAACCTTGAGCAAGCAAAGAAA
ATTAGTTACTCAAACAAACCCCTTGA

30 <210> 2
<211> 1014
<212> DNA
<213> Homo sapiens

35	<400>	2	atgaatgagc cactagacta ttttagcaaat gcttctgatt tcccccgatta tgcatcgct	60			
	tttggaaatt	gcactgatga	aaacatcccc	ctcaagatgc	actacacctcc	tgttatatat	120
	ggcatttatct	tcctcgffff	atttccagge	aatgcagttag	tgtatccac	ttacatttc	180
	aaaatgagac	cttggaaagag	cagcacccatc	attatgcgtga	acctggccctg	cacagatctg	240
40	ctgtatctga	ccagccccc	cttcctgatt	cactactatg	ccagtggcga	aaactggatc	300
	tttggagatt	tcatgtgtaa	gtttagccgc	ttcagcttcc	atttcaacct	gtatagcagc	360
	atccctttcc	tcacctgttt	cagcatcttc	cgctactgtg	tgtatcattca	cccaatgagc	420
	tgcttttcca	ttcacaaaaac	tcgatgtgca	gttgttagct	gtgtgtgtgt	gtggatcatt	480
	tcactggtag	ctgtcattcc	gatgaccctc	ttgatcacat	caaccaacag	gaccaacaga	540
45	tcagcctgtc	tcgacccac	cagttcggtat	gaactcaata	ctattaagtg	gtacaacctg	600
	attttgactg	caactacttt	ctgcctcccc	ttggtgatag	tgacactttg	ctataccacag	660
	attatccaca	ctctgaccca	tggactgcaa	actgacagct	gccttaagca	gaaaggcacga	720
	aggctaacca	ttctgtact	ccttgcattt	tacgtatgtt	ttttaccctt	ccatatcttg	780
	agggtcattc	ggatcgaaatc	tcgcctgttt	tcaatcgat	gttccattga	gaatcgatc	840
	catgaagctt	acatcggttc	tagaccatta	gctgctcgat	acacctttgg	taacctgtta	900
	ctatatgtgg	ttgtcagcga	caactttcag	caggctgtct	gctcaacagt	gagatgc当地	960
50	qtaaqcgggaa	accttgagca	agcaaagaaa	attagttact	caaacaaccc	ttga	1014

SEQ ID NO: 3 peptide sequence translation of above

MNEPLDYLANASDFPDYAAFGNCTDENIPLKMHYLPVIYGIIFLVGFPGNNAVISTYIFKMRPWK
 SSTIIMLNLAETDLYLTSPLFLIHYYASGENWIFGDFMCKFIRFSFHFNLYSSILFLTCFSIFRY
 CVIHPMSCFSIHKTRCAVVACAVVWIISLVAVIPMTFLITSTNRTNRSACLDLTSSDELNTIKWY
 NLILTATTFCPLPLVIVTLCYTTIHTLTHGLQTDSCLKQKARRLTILLLAFYVCFLPFHILRVIR
 5 IESRLLSISCSIENQIHEAYIVSRPLAALNTFGNLLYVVSDNFQQAVCSTVRCKVSGNLEQAKK
 ISYSNNPZ

<210> 3
 <211> 337
 10 <212> PRT.
 <213> Homo sapiens

<400> 3
 Met Asn Glu Pro Leu Asp Tyr Leu Ala Asn Ala Ser Asp Phe Pro Asp
 15 1 5 10 15
 Tyr Ala Ala Ala Phe Gly Asn Cys Thr Asp Glu Asn Ile Pro Leu Lys
 20 20 25 30
 20 Met His Tyr Leu Pro Val Ile Tyr Gly Ile Ile Phe Leu Val Gly Phe
 35 35 40 45
 Pro Gly Asn Ala Val Val Ile Ser Thr Tyr Ile Phe Lys Met Arg Pro
 50 50 55 60
 25 Trp Lys Ser Ser Thr Ile Ile Met Leu Asn Leu Ala Cys Thr Asp Leu
 65 65 70 75 80
 Leu Tyr Leu Thr Ser Leu Pro Phe Leu Ile His Tyr Tyr Ala Ser Gly
 30 85 90 95
 Glu Asn Trp Ile Phe Gly Asp Phe Met Cys Lys Phe Ile Arg Phe Ser
 100 100 105 110
 35 Phe His Phe Asn Leu Tyr Ser Ser Ile Leu Phe Leu Thr Cys Phe Ser
 115 115 120 125
 Ile Phe Arg Tyr Cys Val Ile Ile His Pro Met Ser Cys Phe Ser Ile
 40 130 135 140
 His Lys Thr Arg Cys Ala Val Val Ala Cys Ala Val Val Trp Ile Ile
 145 145 150 155 160
 Ser Leu Val Ala Val Ile Pro Met Thr Phe Leu Ile Thr Ser Thr Asn
 45 165 170 175
 Arg Thr Asn Arg Ser Ala Cys Leu Asp Leu Thr Ser Ser Asp Glu Leu
 180 180 185 190
 50 Asn Thr Ile Lys Trp Tyr Asn Leu Ile Leu Thr Ala Thr Thr Phe Cys
 195 195 200 205
 Leu Pro Leu Val Ile Val Thr Leu Cys Tyr Thr Thr Ile Ile His Thr
 55 210 215 220
 Leu Thr His Gly Leu Gln Thr Asp Ser Cys Leu Lys Gln Lys Ala Arg
 225 230 235 240

	Arg Leu Thr Ile Leu Leu Leu Ala Phe Tyr Val Cys Phe Leu Pro			
	245	250	255	
	Phe His Ile Leu Arg Val Ile Arg Ile Glu Ser Arg Leu Leu Ser Ile			
5	260	265	270	
	Ser Cys Ser Ile Glu Asn Gln Ile His Glu Ala Tyr Ile Val Ser Arg			
	275	280	285	
10	Pro Leu Ala Ala Leu Asn Thr Phe Gly Asn Leu Leu Leu Tyr Val Val			
	290	295	300	
	Val Ser Asp Asn Phe Gln Gln Ala Val Cys Ser Thr Val Arg Cys Lys			
	305	310	315	320
15	Val Ser Gly Asn Leu Glu Gln Ala Lys Lys Ile Ser Tyr Ser Asn Asn			
	325	330	335	

20 Pro

SEQ ID NO: 4 mouse Tarzan ORF

25	ATGATTGAGCCACTGGACAGTCCAGCCAGTGATTGGATTACCAAGTGCTCTGGGA AACTGCACCGACGAGCAAATCTCATTCAAGATGCAGTACCTCCCGTCATCTATAGCATCATCTTC CTCGTGGGCTTCCCGGGAACACAGTGGCCATCTCCATCTACATTTCAGATGCGGCCGTGGAGG GGCAGTACAGTCATCATGCTGAACCTGGCCTGACGGACTGCTGTATCTGACCAGCCTCCGTTG CTCATCCATTACTATGCCAGTGGTAAAAGTGGATCTTGGAGATTTCATGTGCAAGTTCATCCGC TTCGGCTTCACTTCAACCTCTACAGCAGCATTCTCTCCTCACCTGCTTCAGTCTCCGTTAC 30
	GTTGTGATCATTACCGATGAGCTGCTTTCTATTAGAAAACGCTGGCAGTGGTAGCTTGT GCCGGGGTGTGGGTCAATTCTTGGTAGCTGTCATGCCATGACTTTCTGATCACATCAACCACC CGGACCAATAGGTCTGCTGCCCTGACCTCACCAGTTCAGATGACCTCACTACTATCAAGTGGTAC AATCTCATTGACAGCCCACACTTCTGCCTGCCATTGGTAGTGTGACACTTGCTACACGACA ATTATCAGTACCCCTGACTCACGGGCCTCGGACCCACAGGCTGCTTAAGCAGAAGGCTCGGAGACTG 35
	ACTATTCTGCTCCTCCTGTTCTATATATGTTCTACCCCTCCACATCTGAGGGTCATTGG ATCGAATCTCGCCTGCTTCAATCAGCTGCTCCATCGAGAGTCACATCCACGAAGCTTACATTGTT TCTAGACCATTAGCTGCTCTAACACCTTGGCAACCTGCTGTTATATGTTGGTCAGCAATAAC TTCCAGCAGGCATTCTGCTCTATAGTGAGATGCAAAGCCAGTGGGACCTGAAACAAGGAAAGAAA GACAGTTGCTCAAACAACCCTTGA

40

<210> 4
<211> 501

45 <212> DNA
<213> Mus musculus

<400> 4		
atgattgagc cactggacag tccagccagt gattcgatt tcctggatta cccaagtgt	60	
ctggaaact gcaccgacga gcaaatctca ttcaagatgc agtaccttc cgtcatctat	120	
50 agcatcatct tcctcggtgg ctccccggg aacacagtg ccacatccat ctacatttc	180	
aagatgcggc cgtggagggg cagtagtc atcatgctga acctggcctt gacggacttg	240	
ctgtatctga ccagcctccc gttcctcattactatg ccagtggta aaactggatc	300	
tttggagatt tcatgtgcaa gttcattccgc ttccggttcc acttcaacct ctacagcagc	360	
attctcttcc tcacctgttt cagtccttc cgttacgttg tgatcattca cccgatgagc	420	
55 tgctttcta ttccagaaaac tcgctggca gtggtagctt gtgccgggt gtgggtcatt	480	
tcttggtag ctgtcatgcc c	501	

SEQ ID NO: 5 peptide sequence translation of above

5 MIEPLDSPASDSDL DYP SALGN CTDEQISFKMQ YLPVI YSI IFLVGFP GNTVAI SIYIFKMR PWR
 GSTVIMLNLA LDT DLLYL TSLPFLIH YYASGENWIFGDFMCKFIRFGFH FNLYSSILFLT CFSL FRY
 VVIIHPMSCFSIQKTRWAVVACAGVWVISLVAVMPMTFLITSTTRTNRSACLDLTSSDDLTTIKWY
 NLILTATTFCPLVIVTLCYTTIISTLTHGPRTHSCFKQKARRLTILLLVFYICFLPFHILRVIR
 IESRLLSISCSIESHIHEAYIVSRPLAALNTFGNLLYVVVSNNFQQAFCSIVRCKASGDLEQGKK

10 DSCSNNPZ

<210> 5
 <211> 167
 <212> PRT
 15 <213> Mus musculus

<400> 5
 Met Ile Glu Pro Leu Asp Ser Pro Ala Ser Asp Ser Asp Phe Leu Asp
 1 5 10 15

20 Tyr Pro Ser Ala Leu Gly Asn Cys Thr Asp Glu Gln Ile Ser Phe Lys
 20 25 30

25 Met Gln Tyr Leu Pro Val Ile Tyr Ser Ile Ile Phe Leu Val Gly Phe
 35 40 45

Pro Gly Asn Thr Val Ala Ile Ser Ile Tyr Ile Phe Lys Met Arg Pro
 50 55 60

30 Trp Arg Gly Ser Thr Val Ile Met Leu Asn Leu Ala Leu Thr Asp Leu
 65 70 75 80

Leu Tyr Leu Thr Ser Leu Pro Phe Leu Ile His Tyr Tyr Ala Ser Gly
 85 90 95

35 Glu Asn Trp Ile Phe Gly Asp Phe Met Cys Lys Phe Ile Arg Phe Gly
 100 105 110

40 Phe His Phe Asn Leu Tyr Ser Ser Ile Leu Phe Leu Thr Cys Phe Ser
 115 120 125

Leu Phe Arg Tyr Val Val Ile Ile His Pro Met Ser Cys Phe Ser Ile
 130 135 140

45 Gln Lys Thr Arg Trp Ala Val Val Ala Cys Ala Gly Val Trp Val Ile
 145 150 155 160

Ser Leu Val Ala Val Met Pro
 165

50

SEQ ID NO: 6 Fusion protein of TARZAN with V5 and His tags at c-terminal end.

55 MNEPLDYLANASDFPDYAAFGNCTDENIPLKMHYLPVIYGIIFLVGFPGNNAVVI STYIFKMR PWK
 SSTIIMLNLA CT DLLYL TSLPFLIH YYASGENWIFGDFMCKFIRFSFH FNLYSSILFLT CFSI FRY
 CVIIHPMSCFSIHKTRCAVVACAVVWIISLVAVIPMTFLITSTNRTNRSACLDLTSSDELNTIKWY

NLILTATTFCPLVIVTLCYTTI IHTLTHGLQTDCLKQKARRLTILLLAFYVCFLPFHILRVIR
IESRLLSISCSIENQIHEAYIVSRPLAALNTFGNLLLYVVVSDFNQQAVCSTVRCKVSGNLEQAKK
ISYSNNP RAIRELGKPIPNPLLGLDSSHHHHH

5

SEQ ID NO: 7

10 GCTAGCACCATGAATGAGCCACTAGACTATTAGCAAATGCTCTGATTCCCCGATTATGCAGCT
GCTTTGGAAATTGCACTGATGAAAACATCCCCTCAAGATGCACTACCTCCCTGTATTATGGC
ATTATCTTCCTCGTGGGATTCCAGGCAATGCAGTAGTGATATCCACTTACATTTCAAAATGAGA
CCTTGGAAAGAGCAGCACCATCATTATGCTGAACCTGGCCTGCACAGATCTGCTGTACTGACCAGC
CTCCCCCTCCTGATTCACTACTATGCCAGTGGCGAAAACGGATCTTGGAGATTTCATGTGTAAAG
TTTATCCGCTTCAGCTCATTCAACCTGTATAGCAGCATCCTCTCCTCACCTGTTCAGCATTCA
TTCCGCTACTGTGTGATCATTCAACCAATGAGCTGCTTTCCATTCAACAAACTCGATGTGCAGTT
GTAGCCTGTGCTGTGGTGTGGATCATTCACTGGTAGCTGTCACTCCGATGACCTCTTGATCACA
15 TCAACCAACAGGACCAACAGATCAGCCTGTCTGACACTCACCAAGTTCGGATGAACTCAAACTATT
AAGTGGTACAACCTGATTTGACTGCAACTACTTCTGCCTCCCTGGTGATAGTGACACTTTGC
TATACCACGATTATCCACACTCTGACCCATGGACTGCAAACGTACAGCTGCCTTAAGCAGAAAGCA
CGAAGGCTAACCAATTCTGCTACTCCTGCATTACGTATGTTTACCCCTCCATATCTTGAGG
GTCATTGGATCGAATCTGCCTGCTTCAATCAGTTGAGAATCAGATCCATGAAAGCT
20 TACATCGTTCTAGACCATTAGCTGCTCTGAACACCTTGGTAACCTGTTACTATATGTGGTGGTC
AGCGACAACCTTCAGCAGGCTGTGCTCAACAGTGAGATGCAAAGTAAGCGGGAACCTTGAGCAA
GCAAAAGAAATTAGTTACTCAAACAAACCCCTGACTCGAG

SEQ ID NO: 8

25 SEQ ID NO.:
MNEPLDYLANASDFPDYAAAFGNCTDENIPLKMHYPVIYGIIFLVGFPGNAVVISTYIFKMRPWKS
SSTIIMLNIACTDLLYLTSPLPFLIHYYASGENWIFGDFMCKFIRFSFHFLNLYSSILFLTCFSIFRY
CVIIPHMSCFSIHKTRCAVVACAVVWIISLVAVIPMTFLITSTNRTNRSACLDLTSDELNTIKWY
NLILTATTFCPLPVIVTLCYTTIHTLTGLQTDCLKQKARRLTILLLAFYVCFLPFHILRVIR
IESRLLSISCSIENQIHEAYIVSRPLAALNTFGNLLLYVVVDNFQQAVCSTVRCKVSGNLEQAKK
30 ISYSNNPDYKDDDK

SEQ ID NO: 10 genomic locus from 5'prF to 3'prR

GTGGTAGCTTGTGCGGGGTGTTGGTCATTTCTTGGTAGCTGTCATGCCATGACTTCTGATCACATCAACCCACCGGACCAATAGGTCTGCTGCG
 5 CACCATCGAACCGCCCCAACCCAGATAAGAACCCATCGACACTGGGACTCTGAAGGACTAGTGTAGTTGGCTGGCTGGTATTACAGACGACCG
 V V A C A G V H V I S L V A V N P M T P L I T S T T R T N R S A C
 MUSTARZAN
 EXON 4

32900
 TTGACCTCACCACTGAGATGACCTCACTATCAAGTGGTACAATCTCATTTCAGCAGCCCACCTTCTGCTGCCATTGGTGTAGTGACACTTG
 10 ACTGGGACTGGTCAGTCTACTGGAGTGTAGTGTAGTGTAGTAAAGAACCCATCGACACTGGGACTCTGAAGGACTAGTGTAGTTGGCTGGCTGGT
 L D L T S S D D L T T I K W Y N L I L T A T T F C L P L V I V T L C
 MUSTARZAN
 EXON 4

15 CTACACGACATTATCAGTACCCCTGACCTCACGGGCTCGGACCCAGCTGCTTAAAGCAGAAGGCTGGAGACTGACTATTCTGCTCTCCCTGTT
 20 GATGTCGCTTAATAGTCAGTGGACTGAGTGGCCGGAGCTGGTGTGAGAACCTGCTGAGGAAAGTGTGCTGAGGCTCTGACTGATAAGACGAGGAGAACAAAG
 Y T T I S T L T H G P R T H S C F K Q K A R R L T I L L L L V P
 MUSTARZAN
 EXON 4

25 >hetF
 TATATATGTTCTTACCCCTTCCACATCTTGAGGGTCATTGGATCGAATCTCCGCTGTTCAATCAGCTGCTCCATCGAGACTCACATCCACCGAAC
 ATATATACAAAAGATGGGAGGTGAGAAGTCCAGTAAGCTTACGCTTACAGGGACCAAAGTGTAGTCGACGGAGGTAGCTCAGTAGGTAGGCTCTCCGA
 Y I C F L P P H I L R V I R I E S R L L S I S C S I E S H I H E A
 MUSTARZAN
 EXON 4

30 ACATTTGTTCTAGACCAATTAGCTGCTCAACACCCTGGCACCTGGTGTATATGTTCTGGTCAGCAATACTTCAGCAGGCCATTCTGCTCTCTAGT
 35 TGTAAACAGATCTGGTAATCGAGGAGTGTGGAAACCTGCTGAGCAGAACATACACAGGAGTCGTTATIGAAGCTGCTCGTAAAGGAGATCTCA
 Y I V S R P L A A L N T F G N L T V V V S N N F Q Q A F C S I V
 MUSTARZAN
 EXON 4

40 >3' arm
 >3' armP <hetR
 33300
 GAGATGCAAGGCCAGTGGGGACCTTGAACAGGAAGAAGAGACAGTGGCTCAACAAACCCCTGAGTAATGCCATTGGTCAAGGCCAGAGATGGC
 45 CTCTACGGTTGGTCACTGGACTCTGCTTCTCTGTCACCGAGTTGGAAACTCTACGGTAAAGCAGTCTCGGTCTCTACCGG
 R C K A S G D L E O G K R D S C S N N P >
 MUSTARZAN
 EXON 4

50 ACTGGTAAAGCTCTGACATGAAATGAGGGCTGAGATAGACTCTCAGAGGCCACATAAATGGTGTGGTAGCAACTATCTACAACTCCAGTG
 55 TCACCATTCGAGAACTCTACGTTACCTCCAGACTCTACGGTGTGTTAAACACACCATCTGGTGTAGATGTTAGGTCAGAGATGGCTCAAG
 AGACACCACTCTATCTATACCTCTATCTAGGAACTTCCGAGCTTCCGAGTGCCTGGAGTGTGTACTGAGGTTCTCTGGAGAAAGGGT
 60 33600
 GAGATAAAAGGGAGGACCAATGCTTAAAGGCTGCTTGTAGCTACAGTCTACGAGCTGATGACATGACATGACATAAACAGCCACGCATCA
 CTCTATTTCTCTCTCTGGTACGAACTCCGACAGGAAACTGGAGATGTACAGTGTACATGCTGGACTGTATTTGTGGTGGCTAGTAG
 65 33700
 ACACATACATACTTATCAAATATCAGTCTTCCATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA
 TGTGTATGTATGAAATGTTATAGTCGAAAGGTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGGCT
 70 33800
 ATCTCTGAATACCCCTCTAGTGGAGGCCACAGTAACACTGAGTATTCTGGCTTATGAAACCAAGCCATGGTAAGGGCCATTAGAGATCTCTGGCAAT
 TAGGACTTATGGGAGAGTCACCTCGGGTGTCTGGTACGATGACTCATAACTCTGGTACCCATTCCGGAAATACTCTAGGATACCGT
 75 33900
 CCAACTTAATGCCCTCACACAGAAACTAGTGGCTCGTACACTAAACCCGGACCTCTACAGCTTAACTCCAGATGGGAGGCTCAAACAAACGAC
 GTGGTATTAAAGGGAGGTGTGCTTCTGGTACGAGCTGTGTTGGGGCTTGGAGCTGCAATTACGGTTCTCCGAGTTGGTCTGCT
 80 34100
 CAAGCAAAGCCAAGACACAGAAATATGGCTGAGGGAGATCTACGGCTTCTCCACCCAGGAAGTGAATATTAGTGGAGATAACTCTGGCTG
 GTCTGTTTCTGGTTCTGCTTCTGCTTCTACCGACTCGGTAGTGGCAAGACGGTGTCTCTACCTTAACTCCTGCTTCTGGAGGCT
 85 34200
 GATGCGATGGAATCTGGAGATCAAACCAAGTATPCCAACTGGTTCTGTGATCATGGAAAAAGTCTCTGGCTCTGAGTTGTGACACTGAGAAAG
 CTACGTTACCTTACGACCTCTCTAGTTCTGTTCTATAGGTGGACCAAGGACACTAGTACCTTTCTAGAGACGCCAGAGTCTAACAACTGCT
 90 34300
 GAAGTACTTGGCCCATCTGACATCACCTTGGCTGAGTCTCTGGATGATCAGGAAAGCCTCCAGCTGCTCTGGAGAAGGCCACCTCTGAT
 CTCTCATGAAACCCAGGTAGACTGTAGTGGACCCAGTCAGGACCTACTAGTCTTCTGAGTGCAGCAGGAGCTTCTGGTGGAGGACTAAC
 95 34400
 ACAGGACAGTCCCACAGTCACCTAATGAGATAGCTTACAAAGGAAAAGGAGAAAGTGGATTCTGTATGTGGCTTACATGGGAGGAGAAC
 TOTCTCTGTCAGGGTGTGAGTGTGTTAGTACTCTATGAAATGTTCTCTTCTCTACCTAAAGACATACACGGATGTGAGGCTCTGGAG
 100 34500
 TTTCGGCTCTGGGGCCCTTGCTGGGCTAGGGCTCACCCCTCTGCTCTGTTCTGTTCTGTTAACTGGCTGTTCTGCTTACTCCGGAGGCT
 AAACCCAGAACACTCCGGAAACGACCCGATCGAGTGGGAAGGAGCAACATAGAATTTAGCCTGAGACTGAGGCTCTGGCT
 105 34600
 CGACACTGCTAGTCGATTTGTAATGGATTTTCTCTACTATATGGAAAGGTTTCTCTTCTGTTAGTGTAGTGTGGCTTACATGGGAGGAAAG
 GTCTGGAGGAGTCACTGAAACCTGACTTACGTTCTGGCTGTTCTGCTACTGGNATCTAGTATTCACAGGCTCTGGCT
 110 34700
 TTCTCTCATGCTCTAGTTCTGCTTCTGAGCTTCTGGGCTTGTGCTACTGGNATCTAGTATTCACAGGCTCTGGCT
 AAAGAAGTACGGATCAAAGGACACGAAGTCGAAACCCGAACACAGTCAGTGGCTTCTGAGTCTAATAGGTTCTGGTAAAGACAC
 34800
 CAACTCATATCTTACAAAGTTTCTCTTATGGCTGAGTTCTCTCTGAGCTCTGGCTGAGCTTCTAGCTGAGGAAATACATTAGT
 GTTGGAGTAATGAAATGTCAAAAGAAGGAAATAACCGACTCAAAGGAAAGTGGAGACCCGATCGAGAAATCTGACTCTCTTGT
 34900
 TGAGAGAAGAGACTCTGGAAATGCGGAAATGCGGAAATGCGGAAATGCGGAAATGCGGAAATGCGGAAATGCGGAAATGCGGAAATGCGGAA

100

AGTCTTCTTCTGACACCTTTACGGGTCAAGTGTGTTACGGACGATATGTTGTA
 35000
 TCGAACAGACATGCCACTTGAAACCCACGGCTCGGGCTGGCAGGACGGCCCTGGAGGGTGAGGCCAGCCAGCTGAGACTACTTATGAGCTCTA
 ACGTTCTGTAGCTGAACCTTGGGTCAAGACCCGACCGTCTCGTCGGACCTCCACTCCGGTCTGGTCGGATCTGATGAATCACTCGAGAT
 35100
 ACACAAATGAAACCCCTGACTCCAAAATGAGCAACCACAGCAAGGGTGGCCCTGCGCTGAGTTTCAGCCTTGAGGCCACCTTAACAATTGCTTAAT
 TGTTGTTATCTTGGGACTGAGCTTTCAGCTGCTGGTGTCTCCACGGGACCGGACTCAAAGTCGAAACATCGGAAATGTTAACGAATTA
 35200
 TTATAATCATATTAAGTTTATGGTAATGTCACATTTCTGGGAAAGGAAATTTGGTAATAGAGCTACTGTTAGTATCATATATACATTAGGGC
 ATATTTAGTATAATTCAAAATACCCATTAAACAGTTGAAAGGACCCCTTCCTTAAACCCACTTATCTCGATGACCAATCATAGTATATGTTAACCGC
 35300
 CTAACTGAAATCTAACATTGTTTGATTTCTGGGATAGATCTCAGGAAACACTTATCTCTTGTAGTATGTTAATCTTGTAGTACTTAAATAAAAAA
 GATTGAGTTAGATTGAAACAAACTAAAGTACCCCTACTAGGTCCTTGTGATAAAGGAAATCCATAACATTAGGAATCATGATATTATTTT
 35400
 TAATACTGAGAAATGCTAAAGTATCTTCTGGGCTATGTTGCTTTAGTAGTAAATAAGTCAGTGTGATAATAATTTGAGGAAATGCATGG
 ATTATTGACATCTTAAACGATTCTAAAGAACCCGATACAAACGGATAATCTCATTATTCAGTCAACTATATTATAACCTCTTACCTTAC
 35500
 TCTAGATCTTAAACCTTAAAGAATGTTGATGACTAAATAACCTACATACAACTACTCTCTGAATTGGTCACTTGTCTCCACGAAACTGCGCTT
 AGATCTGAAATTTGGAAATTCCTTACACAATACTGTTTAAATGAGGAGACTTAAACAGTAAACAGGGTGTCTGACCCGAGAA
 35600
 ATTCTTATGTTGTTGAAACATGGTGGAGGACAAAAGGAGCTTGAAGGAAATTCAGCAGTCGAATAATGGTCGGGCTCAGGGCTCTCATAGGAAAG
 TAAGTAATCACATACTTTTGTGACCAAGCTCTGTTTCTCTGGAACCTTTTACTGTCAGCAGCTTAAACCCGAGTCAGGAAATGCTTAC
 35700
 CGATTATCAGACTCTGTGTTCTGACACTTCCGCCCCCTGTCCTGGGAGACAGAAATCCGGAATGCTGAGACTGCAATTACCCCTATGCCCTGTCAGG
 GCTAACTGCTGAGACACCAAGTCGTGAAAGGGGAGACAGACCCCTGCTTAAAGCCTTAAAGCTGATGCTAAATGGGATACGGGAGACGTCCAGGA
 35800
 AGGCTGTGACAAAGGCCACGACCAACTGAGGATCTGGGAAATTCGCAACCTAAAGATGTTGGGAGACCAAAAGTCACAGAGGATTTGGAAATCGGG
 TCCGCACTGTTCTCGGCTCTGTTGACTCTAGACCCCTTAACTCCCTGTTACCCCTCTGGTTACAGCTGCTCTCTAAACCTTGTAGCCCGA
 35900
 CCACGGACCTGGCGTGAGGAATTGAGGCTGGGGCTCTCAAACTCTGGGATCCAGGCGCACTGGGCTCTGGGAGTTGGGAATGGAGTTGGTC
 GGTCCTGAGACCGACTCTTAAACTCTGAGACCCCGAGATTGAGACCCGAGTTGAGACCCGAGTTGGCCTGGGAGACGACCTCAACCCCTAACCC
 36000
 TAAGACCTGAGGAGGACTTACATGCACTCTGAGATGAGGAGCTCCATGCTTACACTGTTTCTGAGTAAAGGAGACTCTCTGAGAGGAGAGGG
 ATTCTGACTCTCCGACCCCTGGTGGCTCTTAACTGAGGAGACTCTGAGTAAAGGAGACTCTCTGAGGAGACTCTCTCTCTCTCTCTCTCT
 36100
 TCTTCTGAGGAGGACTTACATGCACTCTGAGATGAGGAGCTCCATGCTTACACTGTTTCTGAGTAAAGGAGACTCTCTGAGGAGACTCTCTCT
 AGAAAGACTCCCTGAGATCTGAGACATCTACTCCGAGAGACTCCAGGAGACTCCAGGAGACTTACACTTATGCAACACCTCTCTCTCTCT
 36200
 GTTACTGTTTATTCATGTTGAAATTCGGGCTACCCACCTCTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACTT
 CAATGAGACTAAAGTACGACCCCTTAACTCCAGTATGTTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACTT
 36300
 ACTTACTGGCTAACTCTGGGGGCTCTGAGATCATTAGTATGGAGACTCTTCTGAGTCTGAGGACCATAGTCTGAGTCTGAGGAGACTTCTG
 TGAATGACCGATTGAGACCCCGAGATCTAGTAAATCTACCTCTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACT
 36400
 ATGTGACAGAAAGCTGCGCGCGAC
 TACACTGCTCTGAGACACCGCGCGCGTGCT
 36500
 CCCACCAAAAGTACGGGATATAACATCTT
 GGGTGTTTCTGCT
 36600
 TCCACACCTTAACTCCACTCTCCAAATTCTTAACTTTTAACTTTTAACTACCTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACTTAC
 AGGTTGTAATTAGGGTGAAGGAGGTTAAATTTGAAATTAATGGGACCTTAACTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACTT
 36700
 CACCTCGCTTGTATTCACCTTCTGACTCTGAGCTGCTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACTTAC
 GTGGAGGAGGAACTAAGTCGAGAAACAGTACGAGACATCGACTGGACATAGGGGGGGCTTGTGAGTACTGTCAGACGGAGGTTGGGAGAC
 36800
 TTGAGCATAATCACAAAGGTCCAAGCGAAATTGAAATTGAAATTCAGGAAATGTTGTTTATTTCTTAAAGGAGACTTACCTGAGGAGACTTAC
 AACTCTGTTAGTTCTCAGGTTGCTTTAACTTAAAGGTTTACAGCCAAAGCTGATTAACCGGATCTCCAACTGAA
 36900
 TTGAACTGAAATCCCCTCCCTCCAAAGGACAGATAATGATACGGGACTAGTGGAGATCTACAGGAACTTCTTACAGACATCCACGGGAGAGT
 AACTTGACTTACGGGAGGGAGGTTTCTGCTTAACTACTGCGGAGTCTGCTTACGTTGAGTACTGTCAGACGGAGGTTGGGGAGACTTAC
 37000
 AGATAAGAACAAAAGAACGGGATCAAGCTGTTGTTATACAGGAAATGCTGCTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACTTAC
 TCTTACCTGTTCTTCTTCTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACTTACCTGAGGAGACTTACCTGAG
 37100
 CACCTGAGAAATCCAGATAAAAGAACCTCTGAAATGCTGTTCTTAACTACAAATACTCCCAATTGAGTGTGAGTACTGCTCACAGAAATGTT
 GTGGACTCTTCTGAGTCTTAACTGAGCTTACGGGAAAGAAATATGTTACAGGGGTTAAACTACATCACAGGAGTCTGTTACAGGAGTCT
 37200
 AACCCACAGTTAAAGAACATCAGAACACGGGATGTTGTTACCTTAAACATACATGTCACAGGGGTTAAATGTTGTTGAGTGTGTTGTT
 TTGGGAGTCTTCTGAGTCTTAACTGAGCTTACGGGAAAGAAATATGTTACAGGGGTTAAACTACATCACAGGAGTCTGTTACAGGAG
 AACAGAATACAGGAG

- Each of the applications and patents mentioned in this document, and each document cited or referenced in each of the above applications and patents, including during the prosecution of each of the applications and patents ("application cited documents") and any manufacturer's instructions or catalogues for any products cited or 5 mentioned in each of the applications and patents and in any of the application cited documents, are hereby incorporated herein by reference. Furthermore, all documents cited in this text, and all documents cited or referenced in documents cited in this text, and any manufacturer's instructions or catalogues for any products cited or mentioned in this text, are hereby incorporated herein by reference.
- 10 Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications 15 of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the claims.

CLAIMS

1. A Tarzan GPCR polypeptide comprising the amino acid sequence shown in SEQ ID NO. 3 or SEQ ID NO: 5, or a homologue, variant or derivative thereof.
2. A nucleic acid encoding a polypeptide according to Claim 1.
- 5 3. A nucleic acid according to Claim 2, comprising the nucleic acid sequence shown in SEQ ID No. 1, SEQ ID No.2 or SEQ ID NO: 4, or a homologue, variant or derivative thereof.
4. A polypeptide comprising a fragment of a polypeptide according to Claim 1.
5. A polypeptide according to Claim 3 which comprises one or more regions which
10 are homologous between SEQ ID No. 3 and SEQ ID No. 5, or which comprises one or more regions which are heterologous between SEQ ID No. 3 and SEQ ID No. 5.
6. A nucleic acid encoding a polypeptide according to Claim 4 or 5.
7. A vector comprising a nucleic acid according to Claim 2, 3, or 6.
8. A host cell comprising a nucleic acid according to Claim 2, 3, or 6, or vector
15 according to Claim 7.
9. A transgenic non-human animal comprising a nucleic acid according to Claim 2, 3 or 6, or a vector according to Claim 7.
10. A transgenic non-human animal according to Claim 9 which is a mouse.

11. Use of a polypeptide according to Claim 1, 4 or 5 in a method of identifying a compound which is capable of interacting specifically with a G protein coupled receptor.
12. Use of a transgenic non-human animal according to Claim 9 or 10 in a method of identifying a compound which is capable of interacting specifically with a G protein coupled receptor.
13. A method for identifying an antagonist of a Tarzan GPCR, the method comprising contacting a cell which expresses Tarzan receptor with a candidate compound and determining whether the level of cyclic AMP (cAMP) in the cell is lowered as a result of said contacting.
- 10 14. A method for identifying a compound capable of lowering the endogenous level of cyclic AMP in a cell which method comprises contacting a cell which expresses a Tarzan GPCR with a candidate compound and determining whether the level of cyclic AMP (cAMP) in the cell is lowered as a result of said contacting.
- 15 15. A method of identifying a compound capable of binding to a Tarzan GPCR polypeptide, the method comprising contacting a Tarzan GPCR polypeptide with a candidate compound and determining whether the candidate compound binds to the Tarzan GPCR polypeptide.
16. A compound identified by a method according to any of Claims 11 to 15.
17. A compound capable of binding specifically to a polypeptide according to Claim 1, 20 4 or 5.
18. Use of a polypeptide according to Claim 1, 4 or 5, or part thereof or a nucleic acid according to Claim 2, 3 or 6, in a method for producing antibodies.

19. An antibody capable of binding specifically to a polypeptide according to Claim 1, 4 or 5, or part thereof or a polypeptide encoded by a nucleotide according to Claim 2, 3 or 6, or part thereof.
20. A pharmaceutical composition comprising any one or more of the following: a 5 polypeptide according to Claim 1, 4 or 5, or part thereof; a nucleic acid according to Claim 2, 3 or 6, or part thereof; a vector according to Claim 7; a cell according to Claim 8; a compound according to Claim 16 or 17; and an antibody according to Claim 19, together with a pharmaceutically acceptable carrier or diluent.
21. A vaccine composition comprising any one or more of the following: a polypeptide 10 according to Claim 1, 4 or 5, or part thereof; a nucleic acid according to Claim 2, 3 or 6, or part thereof; a vector according to Claim 7; a cell according to Claim 8; a compound according to Claim 16 or 17; and an antibody according to Claim 19.
22. A diagnostic kit for a disease or susceptibility to a disease comprising any one or 15 more of the following: a polypeptide according to Claim 1, 4 or 5, or part thereof; a nucleic acid according to Claim 2, 3 or 6, or part thereof; a vector according to Claim 7; a cell according to Claim 8; a compound according to Claim 16 or 17; and an antibody according to Claim 19.
23. A method of treating a patient suffering from a disease associated with enhanced 20 activity of a Tarzan GPCR, which method comprises administering to the patient an antagonist of Tarzan GPCR.
24. A method of treating a patient suffering from a disease associated with reduced activity of a Tarzan GPCR, which method comprises administering to the patient an agonist of Tarzan GPCR.
25. A method according to Claim 23 or 24, in which the Tarzan GPCR comprises a 25 polypeptide having the sequence shown in SEQ ID NO: 3 or SEQ ID NO: 5.

26. A method for treating and/or preventing a disease in a patient, which comprises the step of administering any one or more of the following to the patient: a polypeptide according to Claim 1, 4 or 5, or part thereof; a nucleic acid according to Claim 2, 3 or 6, or part thereof; a vector according to Claim 7; a cell according to Claim 8; a compound according to Claim 16 or 17; an antibody according to Claim 19; a pharmaceutical composition according to Claim 20; and a vaccine according to Claim 20.
- 5
27. An agent comprising a polypeptide according to Claim 1, 4 or 5, or part thereof; a nucleic acid according to Claim 2, 3 or 6, or part thereof; a vector according to Claim 7; a cell according to Claim 8; a compound according to Claim 16 or 17; and/or an antibody according to Claim 19, said agent for use in a method of treatment or prophylaxis of disease.
- 10
28. Use of a polypeptide according to Claim 1, 4 or 5, or part thereof; a nucleic acid according to Claim 2, 3 or 6, or part thereof; a vector according to Claim 7; a cell according to Claim 8; a compound according to Claim 16 or 17; and an antibody according to Claim 19, for the preparation of a pharmaceutical composition for the treatment or prophylaxis of a disease.
- 15
29. A non-human transgenic animal, characterised in that the transgenic animal comprises an altered Tarzan gene.
30. A non-human transgenic animal according to Claim 29, in which the alteration is selected from the group consisting of: a deletion of Tarzan, a mutation in Tarzan resulting in loss of function, introduction of an exogenous gene having a nucleotide sequence with targeted or random mutations into Tarzan, introduction of an exogenous gene from another species into Tarzan, and a combination of any of these.
- 20
31. A non-human transgenic animal having a functionally disrupted endogenous Tarzan gene, in which the transgenic animal comprises in its genome and expresses a transgene encoding a heterologous Tarzan protein.
- 25

32. A nucleic acid construct for functionally disrupting a Tarzan gene in a host cell, the nucleic acid construct comprising: (a) a non-homologous replacement portion; (b) a first homology region located upstream of the non-homologous replacement portion, the first homology region having a nucleotide sequence with substantial identity to a first Tarzan gene sequence; and (c) a second homology region located downstream of the non-homologous replacement portion, the second homology region having a nucleotide sequence with substantial identity to a second Tarzan gene sequence, the second Tarzan gene sequence having a location downstream of the first Tarzan gene sequence in a naturally occurring endogenous Tarzan gene.
- 5
- 10 33. A process for producing a Tarzan GPCR polypeptide, the method comprising culturing a host cell according to Claim 8 under conditions in which a nucleic acid encoding a Tarzan GPCR polypeptide is expressed.
- 15 34. A method of detecting the presence of a nucleic acid according to Claim 2, 3 or 6 in a sample, the method comprising contacting the sample with at least one nucleic acid probe which is specific for said nucleic acid and monitoring said sample for the presence of the nucleic acid.
35. A method of detecting the presence of a polypeptide according to Claim 1, 4 or 5 in a sample, the method comprising contacting the sample with an antibody according to Claim 19 and monitoring said sample for the presence of the polypeptide.
- 20 36. A method of diagnosis of a disease or syndrome caused by or associated with increased, decreased or otherwise abnormal expression of Tarzan GPCR, the method comprising the steps of: (a) detecting the level or pattern of expression of Tarzan GPCR in an animal suffering or suspected to be suffering from such a disease; and (b) comparing the level or pattern of expression with that of a normal animal.
- 25 37. A nucleic acid sequence selected from group consisting of musTarzan 5'prF (TTTGATGGAGTAGATTCACTGTCAGAC), musTarzan 5'prR

- (GTCCCAGAAGAGCTTGGGAATCCAAAC), musTarzan 5'armF.2 Age
(tttacggGTTCCTACATTAGCGACTTTCTGGAC), musTarzan 5'armR Not
(aaagcggccgCCAGGAAATCCGAATCACTGGCTGGAC), musTarzan 3'armF Asc
(aaaggcgccGATGCAAAGCCAGTGGGGACCTTGAAC), musTarzan 3'armR Fse
- 5 (tttggccGGcCTGAGCATGCATGGCAGTTCTTG), musTarzan 5'scr.2
(AATCATCCTCTTGCTGGGTTAGAC), musTarzan 3'prF
(CGTGGCACTCAAGTCTGGATGCACCTG), musTarzan 3'prR
(AGACACTAGGACATTAATCCCTGTGAC), musTarzan hetF
(TCCACATCTTGAGGGTCATTGGATCG), musTarzan hetR
- 10 (TTCTTCCTTGTCAAGGTCCCCACTG), musTarzan F1
(GGATTCCTGGATTACCCAAGTGC), musTarzan R1
(CAGATACAGCAAGTCCGTCAAGGC), Asc306
(AATGGCCGCTTTCTGGATTGATCGAC) and DR1
(CATGCCGCCTGCGCCCTATTGATCATG).
- 15 38. A nucleic acid sequence, including a Tarzan nucleic acid sequence, a polypeptide sequence, including a Tarzan polypeptide sequence, a method of treatment, a method of diagnosis, a host cell, vector, transgenic animal, assay, diagnostic kit, vaccine, pharmaceutical composition or agent substantially as hereinbefore described with reference to and as shown in the accompanying drawings.

FIGURE 1

Domain	Start	End	Bits	Evalue	Alignment
7tm_1	50	302	168.60	2.1e-52	Align

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_87	73	170	4.2e-06	Align

[337 residues]

7tm_1 50-302

Alignments of Pfam-A domains to HMMsFormat for fetching alignments to seed [Hypertext linked to swisspfam](#)Alignment of 7tm_1 vs UserSeq/50-302

User Seq	50	*->GMLVilvilrtkklr..tptnifilNLAvADLLfltlppwallyl GN++Vi + ++ k+r+ +++ i +LNLA +DLL+l++lp+ + yy+ GNAVVIVSTYIF--KIRpwKSSTIIMLNACTDILYLTSPLFLIXHYA 94
User Seq	95	vggosedWpfGsdlCkltalvvnnmyaSillLtaISiDRY1AI+hPlryr g e W fG+++Ck+ +++ +n+y+S1+L+ S1 RY++I+hP++ SG—ENWIFGDFMCKFIRSFHFLYSSILFLTCFSIFRVCVLIKPMSCF 142
User Seq	143	rrrtsprAkvvillvWvlallls1PpllfswwktveegngtlnvnvttvC +++ r A+v ++++W++l++ +P+ ++ ++++ ++ +C SIHK-TRCAVVACAVVWISLVAVIPMTFLITSTKRTNR-----SAC 183
User Seq	184	lidipeestasvstwlsyvllstlvgFlPllvilvcYtrIlrtlr... l +++ ++ ++y+l+t +F lPl++ + cYt I++tl ++ LDLTSSDELNT---IKWYNLILTATTFCCLPLVIVTLCYTTIIXTLThgl 229
User Seq	230kaaktllvvvvvFv1CW1Pyfivllldtlc.lsiimsstCel ++++ ++ka +++++++F +C+lP++i+ +++ +1 s +C+ qtdsclrqKARRLTILLAFVCFLPFHILRVIRIESrlL---SISCSI 276
User Seq	277	ervlpallvtlwLayvNsclNPiIY-* e++++ ++v++ La +N N ++Y ENQIKEAYIVSRPLAALNTFGNLLY 302

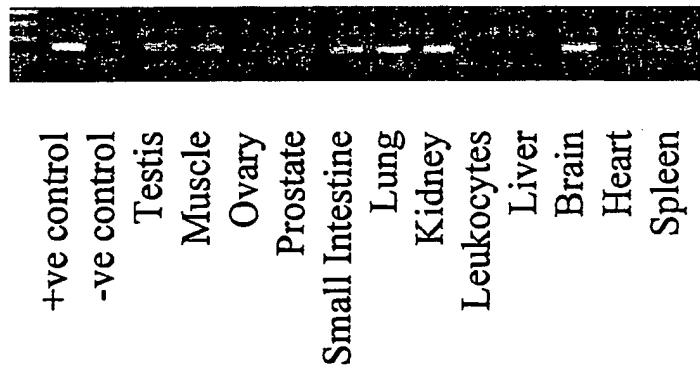
FIGURE 2

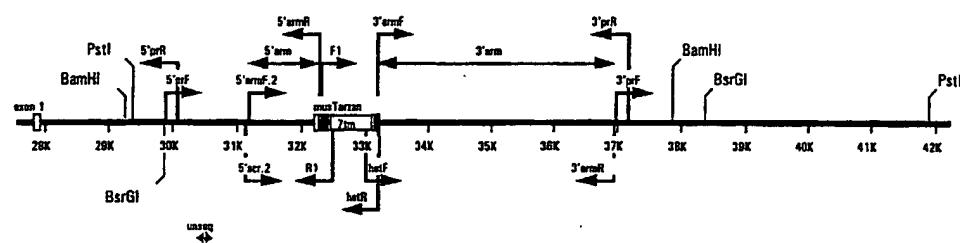
FIGURE 3

FIGURE 4

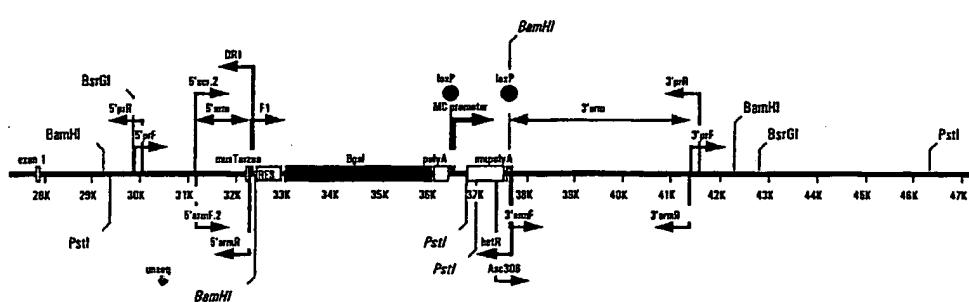
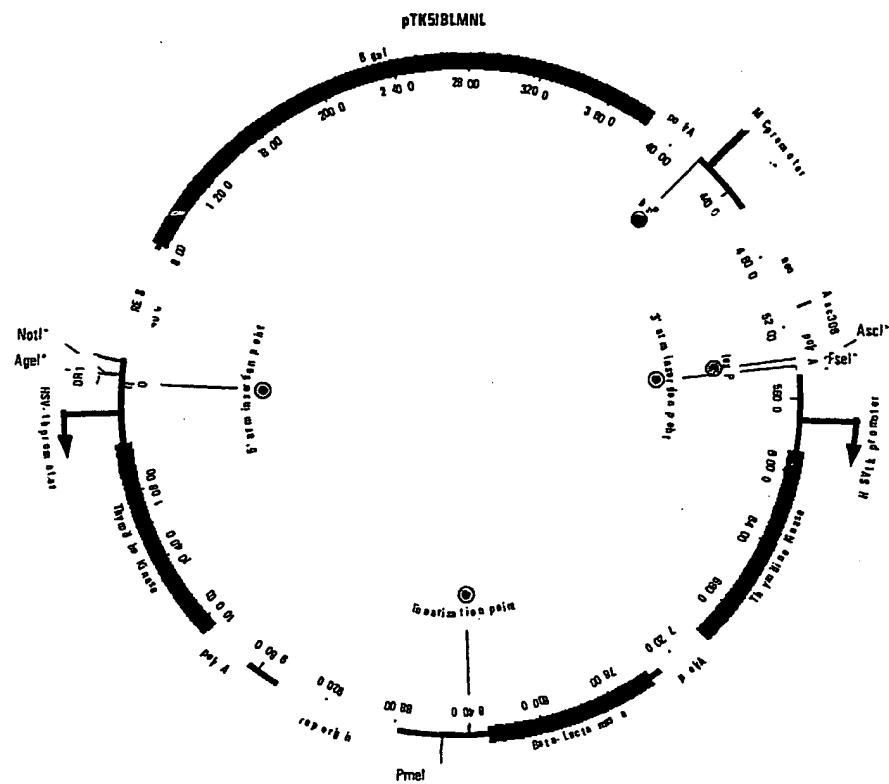


FIGURE 5

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 November 2002 (07.11.2002)

PCT

(10) International Publication Number
WO 02/088183 A3

(51) International Patent Classification⁷: C07K 14/705, 14/72 (74) Agents: KHOO, Chong-Yee et al.; D Young & Co, 21 New Fetter Lane, London EC4A 1DA (GB).

(21) International Application Number: PCT/GB02/01923

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(22) International Filing Date: 25 April 2002 (25.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0110662.4 1 May 2001 (01.05.2001) GB
60/288,160 2 May 2001 (02.05.2001) US

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): PARADIGM THERAPEUTICS LIMITED [GB/GB]; St. John's Innovation Centre, Cowley Road, Cambridge CB4 0WS (GB).

Declaration under Rule 4.17:

— *of inventorship (Rule 4.17(iv)) for US only*

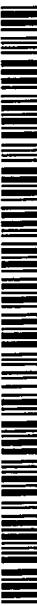
Published:

— *with international search report*

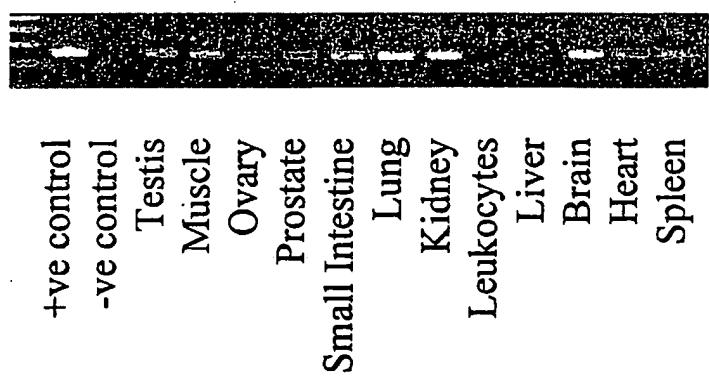
(88) Date of publication of the international search report:
1 May 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: G-PROTEIN COUPLED RECEPTOR



WO 02/088183 A3



(57) Abstract: We disclose Tarzan G-protein coupled receptor (GPCR) polypeptides comprising the amino acid sequence shown in SEQ ID NO: 3, SED ID NO: 5, SEQ ID NO: 6 OR SEQ ID NO: 8, and homologues, variants and derivatives thereof. Nucleic acids capable of encoding Tarzan polypeptide are also disclosed, in particular, those comprising the nucleic acid sequences shown in SEQ ID NO: 1, SEQ ID NO: 2 or SEQ ID NO: 4.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 02/01923

A. CLASSIFICATION OF SUBJECT MATTER
 IPC 7 C07K14/705 C07K14/72

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHEDMinimum documentation searched (classification system followed by classification symbols)
 IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, EMBL, BIOSIS, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL 'Online! 21 May 2000 (2000-05-21) SKUCE C.: "Human DNA sequence from clone RP11-721F14 on chromosome 13" retrieved from EMBL Database accession no. AL356486 XP002216296 abstract --- DATABASE EMBL 'Online! ABOLA ET AL.: "Homo sapiens chromosome 13 clone RP11-286P8, complete sequence." retrieved from EMBL Database accession no. ac026756 XP002216311 abstract --- -/--	1-4
X		1-4

 Further documents are listed in the continuation of box C. Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the International filing date
- "L" document which may throw doubts on priority, claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the International filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the International search	Date of mailing of the International search report
14 October 2002	05.02.03
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel: (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized office: Paresce, D

INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 02/01923

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL 'Online! NCI-CGAP; 18 May 2000 (2000-05-18) ROBERT STRAUSBERG: retrieved from EMBL Database accession no. AW827323 XP002216297 cited in the application abstract ---	1-4
X	TOKUYAMA Y ET AL: "CLONING OF RAT AND MOUSE P2Y PURINOCEPTORS" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, ACADEMIC PRESS INC. ORLANDO, FL, US, vol. 211, no. 1, 6 June 1995 (1995-06-06), pages 211-218, XP000673449 ISSN: 0006-291X see p. 211 and figure 1 ---	1-4, 7, 8, 11, 13-15, 18-20, 33-35
Y	AYYANATHAN KASIRAJAN ET AL: "Cloning and chromosomal localization of the human P2Y1 purinoceptor" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, ACADEMIC PRESS INC. ORLANDO, FL, US, vol. 218, no. 3, 1996, pages 783-788, XP002195807 ISSN: 0006-291X see abstract, p. 784-5 and fig. 2 ---	1-4, 7-15, 18-22, 26-28, 30-36
Y	KUEGELGEN VON I ET AL: "Molecular pharmacology of P2Y-receptors" NAUNYN-SCHMIEDEBERG'S ARCHIVES OF PHARMACOLOGY, SPRINGER, BERLIN, DE, vol. 362, no. 4-5, November 2000 (2000-11), pages 310-323, XP002199630 ISSN: 0028-1298 see p. 311, 319, figures 3, 4 ---	1-4, 7-15, 18-22, 26-28, 30-36
Y	LEE D K ET AL: "Cloning and characterization of additional members of the G protein-coupled receptor family" BIOCHIMICA ET BIOPHYSICA ACTA. GENE STRUCTURE AND EXPRESSION, ELSEVIER, AMSTERDAM, NL, vol. 1490, no. 3, 29 February 2000 (2000-02-29), pages 311-323, XP004275600 ISSN: 0167-4781 see fig. 1-3, p. 312-4 ---	1-4, 7-15, 18-22, 26-28, 30-36

-/-

INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 02/01923

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	WO 01 87980 A (APPLERA CORP) 22 November 2001 (2001-11-22) see Claim 23; Fig 3 and Claim 1; Fig 2; ---	1-4, 7-15, 18-22, 26-28, 30-36
P, X	WO 01 87937 A (INCYTE GENOMICS INC ;KALLICK DEBORAH A (US); PATTERSON CHANDRA (US) 22 November 2001 (2001-11-22) see seq. 2 ---	1-4, 7-15, 18-22, 26-28, 30-36
P, X	WO 01 36471 A (ARENA PHARMACEUTICALS INC ;CHEN RUOPING (US); DANG HUONG T (US); L) 25 May 2001 (2001-05-25) see sequences 27, 28 ---	1-4, 7-15, 18-22, 26-28, 30-36

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB 02/01923

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
see FURTHER INFORMATION sheet PCT/ISA/210
2. Claims Nos.: 5-6, 16-17, 23-25, 29, 38 and 20-22, 26-28 partially because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-36, 38

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
 No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1: claims 1-36, 38

A Tarzan G protein coupled receptor, nucleic acid encoding said receptor and uses thereof.

Inventions 2-16: claim 37

Nucleic acid sequences comprising fragments of a G protein-coupled receptor sequence. Each sequence is considered a separate invention, the inventions are numbered according to the list of sequences given in claim 37, namely, invention 2 is directed to the sequence musTarzan5'prF(TTGATGGA...), invention 3 is directed to the sequence musTarzan5'prR(GTCCCAG...) etc.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claims 12-15, 36 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition. Although claims 23-26 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Continuation of Box I.2

Claims Nos.: 5-6, 16-17, 23-25, 29, 38 and 20-22, 26-28 partially

Present claims 5-6, 16-17, 23-25, 29, 38 and 20-22, 26-28 partially relate to an extremely large number of possible compounds/products/methods. In fact, the claims contain so many options, variables, possible permutations and provisos that a lack of clarity (and/or conciseness) within the meaning of Article 6 PCT arises to such an extent as to render a meaningful search of the claims impossible. Furthermore, claims 23-25 relate to a product/compound/method defined by reference to a desirable characteristic or property, namely that of being an agonist or antagonist of the claimed polypeptide. The claims cover all products/compounds/methods having this characteristic or property, whereas the application provides support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for only a very limited number of such products/compounds/methods. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the product/compound/method by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, the search has been carried out for those parts of the application which do appear to be clear (and/or concise), namely claims 1-4, 7-15, 18-19, 30-36 completely and 20-22, 26-28 partially.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International Application No
PCT/GB 02/01923

Patent document cited in search report	Publication date		Patent family member(s)	Publication date
WO 0187980	A 22-11-2001	AU WO	6171501 A 0187980 A2	26-11-2001 22-11-2001
WO 0187937	A 22-11-2001	AU WO AU WO	6472101 A 0187937 A2 6181401 A 0190359 A2	26-11-2001 22-11-2001 03-12-2001 29-11-2001
WO 0136471	A 25-05-2001	AU CN EP WO	1769601 A 1391581 T 1242448 A2 0136471 A2	30-05-2001 15-01-2003 25-09-2002 25-05-2001